St. Petersburg State University

Graduate School of Management

Master in Management Program

MATURITY OF RUSSIAN ENTERPRISES IN IMPLEMENTATION OF DATA ANALYTICS

Master's Thesis by the 2nd year student

Concentration - Management

Denis V. Egorenko

Research advisor:

Associate Professor

Sergey A. Yablonsky

St. Petersburg

2021

ЗАЯВЛЕНИЕ О САМОСТОЯТЕЛЬНОМ ХАРАКТЕРЕ ВЫПОЛНЕНИЯ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

Я, Егоренко Денис Владиславович, студент второго курса магистратуры направления «Менеджмент», заявляю, что в моей магистерской диссертации на тему «Зрелость Российских предприятий во внедрении аналитики данных», представленной в службу обеспечения программ магистратуры для последующей передачи в государственную аттестационную комиссию для публичной защиты, не содержится элементов плагиата.

Все прямые заимствования из печатных и электронных источников, а также из защищенных ранее выпускных квалификационных работ, кандидатских и докторских диссертаций имеют соответствующие ссылки.

Мне известно содержание п. 9.7.1 Правил обучения по основным образовательным программам высшего и среднего профессионального образования в СПбГУ о том, что «ВКР выполняется индивидуально каждым студентом под руководством назначенного ему научного руководителя», и п. 51 Устава федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет» о том, что «студент подлежит отчислению из Санкт-Петербургского университета за представление курсовой или выпускной квалификационной работы, выполненной другим лицом (лицами)».

(Подпись студента) ______________________

(Дата) 05.06.2021

STATEMENT ABOUT THE INDEPENDENT CHARACTER OF THE MASTER THESIS

I, Denis Egorenko, second year master student, master program «Management», state that my master thesis on the topic "Maturity of Russian Enterprises in Implementation of Data Analytics", which is presented to the Master Office to be submitted to the Official Defense Committee for the public defense, does not contain any elements of plagiarism.

All direct borrowings from printed and electronic sources, as well as from master theses, PhD and doctorate theses which were defended earlier, have appropriate references.

I am aware that according to paragraph 9.7.1. of Guidelines for instruction in major curriculum programs of higher and secondary professional education at St.Petersburg University «A master thesis must be completed by each of the degree candidates individually under the supervision of his or her advisor», and according to paragraph 51 of Charter of the Federal State Institution of Higher Professional Education Saint-Petersburg State University «a student can be expelled from St. Petersburg University for submitting of the course or graduation qualification work developed by other person (persons)».

(Student's signature) ______________________

(Date) 05.06.2021

АННОТАЦИЯ

Автор	Егоренко Денис Владиславович
Тема магистерской диссертации	Зрелость российских предприятий во внедрении аналитики данных
Направление	Менеджмент (программа Master in Management)
Год	2021
Научный руководитель	Яблонский Сергей Александрович
Цель исследования	Понять, подходят ли существующие модели аналитической зрелости для использования в российских предприятиях, и исследовать текущий уровень аналитической зрелости Российских предприятий.
Вопросы исследования	**RQ1** Как российские компании в настоящее время используют аналитику данных? **RQ2** Можно ли считать существующие модели аналитической зрелости применимыми для российских компаний? **RQ3** Какие уровни зрелости аналитики можно определить для российских компаний? **RQ4** Каковы характеристики компаний на каждом уровне зрелости? **RQ5** Что нужно сделать российским компаниям, чтобы выйти на следующий уровень зрелости?
Методология	В исследовании использовался структурированный опрос мнений специалистов в области аналитики данных, работающих в Российских компаниях о том, как в их компании применяют аналитику данных. С последующим применением трех видов анализа полученных результатов. Вначале результаты были описаны с помощью описательного анализа. Затем была применена обратная пошаговая регрессия для подтверждения или опровержения выдвинутых гипотез. На последнем этапе был применен кластерный анализ для выявления уровней аналитической зрелости российских компаний и особенностей, присущих компаниям на каждом уровне.
Основные результаты	Было описано, как российские компании используют аналитику данных, на основании мнений сотрудников, работающих в этих компаниях. В результате проверки гипотез было выявлено, что российские компании используют аналитику данных в соответствии с существующей теорией об использовании аналитики данных. Это говорит о том, что российские компании могут применять существующие модели аналитической зрелости. В результате кластерного анализа были определены уровни аналитической зрелости российских компаний. Было дано описание того, как компании на определенных уровнях аналитической зрелости применяют анализ данных. Были сформулированы рекомендации по развитию аналитики данных для компании на каждом уровне.
Ключевые слова	Аналитика данных, Модель зрелости, Аналитическая зрелость

ABSTRACT

Name of the Master's student	Denis Egorenko
Master thesis title	Maturity of Russian enterprises in implementation of data analytics
Main field of study	Management (Master in Management program)
Year	2021
Academic advisor’s name	Sergey A. Yablonsky
Research goal	To understand whether the existing models of analytical maturity are suitable for use in Russian enterprises, and to investigate the current level of analytical maturity of Russian enterprises.
Research questions	**RQ1** How do Russian companies currently use data analytics? **RQ2** Can the existing models of analytical maturity be considered applicable for Russian companies? **RQ3** What levels of analytics maturity can be determined for Russian companies? **RQ4** What are the characteristics of companies at each maturity level? **RQ5** What Russian companies need to do to get to the next level of maturity?
Methodology	The study used a structured survey of opinions of data analytics specialists working in Russian companies on how their companies use data analytics. With the subsequent application of three types of analysis of the results obtained. Initially, the results were described using descriptive analysis. The backward stepwise regression analysis was then applied to confirm or refute the hypotheses put forward. At the last stage, cluster analysis was applied to identify the levels of analytical maturity of Russian companies and the characteristics inherent in companies at each level.
Main results	It was described how Russian companies use data analytics, based on the opinions of employees working in these companies. As a result of testing the hypotheses, it was revealed that Russian companies use data analytics in accordance with the existing theory about the use of data analytics. This suggests that Russian companies can apply existing models of analytical maturity. As a result of the cluster analysis, the levels of analytical maturity of Russian companies were determined. A description was given of how companies at the identified levels of analytical maturity apply data analytics. Recommendations were formulated for the development of data analytics for the company at each level.
Keywords	Data analytics, Maturity model, Analytical maturity

TABLE OF CONTENTS

INTRODUCTION

In recent decades, due to the development of technology, companies have been able to collect more and more data. The analysis of the collected data allows the company to gain significant advantages, gain new knowledge about the company's processes, products, and the industry in which they work. However, effective implementation of data analytics in a company is not always an easy process. (Morabito, 2015, Posavec & Krajnovic, 2016) Companies are faced with constantly evolving data analysis technologies, and organizational issues that affect the effectiveness of data analytics. So among such problems, there is often a lack of understanding of how analytics can improve business processes, a lack of analytical skills, and a lack of attention from the company's top management to the problems of data analytics. Therefore, in order for companies to be more effective in the field of data analytics, they need to constantly analyze the current state of company data analytics and make appropriate decisions to improve it.

In order to help companies in the process of implementing data analytics, tools such as company analytical maturity models have been developed. There are many different models of analytical maturity. These tools allow to evaluate how effectively a company implements data analytics and how far it has progressed in this area. These tools also allow companies to understand what needs to be done to become more mature in data analytics, therefore, they are a decision-making tool. However, the literature raises several important questions regarding the development and use of analytical maturity models. Many authors agree that due to the rapid development of technologies, the model of analytical maturity requires constant improvement, taking into account the emergence of new technologies. Also, several authors raise the issue of the need to adapt the model of analytical maturity to the conditions of the market in which the company operates. At the moment, most of the analytical maturity models have been developed in a developed market, while no analytical maturity models have been found developed and adapted for emerging markets, such as the Russian market. Thus, the use of existing models of analytical maturity for Russian companies raises doubts about its effectiveness on these markets. Therefore, further research is needed on the possibility of applying existing analytical maturity models for Russian companies. Also in the future, an important research issue is the development of an analytical maturity model developed and adapted for the Russian market.

The identified problems in the field of analytical maturity models became the basis for this study. The purpose of this study is to understand whether it is possible to apply existing models of analytical maturity in Russian enterprises, and to study how Russian companies currently use data analytics and what level of analytical maturity they have. This will give an understanding of whether Russian companies can use the popular models of analytical maturity that already exist and will lay the foundation for further development of an analytical maturity model adapted specifically for the conditions of the Russian market and Russian companies. With this in mind, the research goal and research questions were formulated.

The research goal of the thesis is to understand whether the existing models of analytical maturity are suitable for use in Russian enterprises, and to investigate the current situation with the analytical maturity of Russian enterprises.

The **research questions** can be formulated as follows:

RQ1 How do Russian companies currently use data analytics?

RQ2 Can the existing models of analytical maturity be considered applicable for Russian companies?

RQ3 What levels of analytics maturity can be determined for Russian companies?

RQ4 What are the characteristics of companies at each maturity level?

RQ5 What Russian companies need to do to get to the next level of maturity?

As a result of the study, all the research questions were answered. It was investigated how Russian enterprises currently use data analytics. The study also confirmed that Russian companies have no major differences in the use of data analytics from foreign companies. This may indicate that the existing models of analytical maturity can be effective for Russian companies. Also, the companies studied were divided into four levels of analytical maturity. The characteristics of companies at each of the revealed levels of analytical maturity were described. Recommendations were also formulated for the company at each of the four levels of analytical maturity, on what companies need to focus on to move to the next level of analytical maturity. The results obtained are the basis for future research in the development of an analytical maturity model adapted for Russian companies. The results also have practical managerial applications, as they offer concrete actions for companies to improve data analytics.

CHAPTER 1. THEORETICAL BACKGROUND

1.1. Big Data

The term Big Data was first mentioned by Roger Magoulas from O'Reilly media in 2005. (Moed, 2012) Since then, Big data has been a trending topic in both the business and scientific environment. The international consulting agency McKinsey in 2011 defined big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze". (J. Manyika et al. 2011) Indeed, the amount of data generated has increased significantly over the past decade due to the advent of new technologies. Thus, according to the Statista website, the volume of data created, captured, copied and consumed in 2010 was only 2 zettabytes, while in 2020 the volume of such data was already 59 zettabytes. That is, it has increased almost 30 times. It is estimated that the volume of processed data will increase to 149 zettabytes in 2024. (Holst, 2021) This significant increase in the volume of data collected and processed due to the development of cloud technologies and IoT has led to an increase in the popularity of the term Big Data and a great interest in this area. It is believed that working with Big data opens up huge business opportunities, but it also causes a lot of problems.

Despite the great popularity of the term big data, there is a debate about the definition of what can be considered big data and what characteristics it has. There are several models that define what big data is and what characteristics it has. The first model describing the challenges and opportunities of increased data volumes was proposed in 2001 by Doug Laney, an analyst at META/ Gartner/ *now Data & Analytics Innovation Fellow at West Monroe.* (Chen, Mao, & Liu, 2014) Later, with the advent of the term Big Data, this model was widely used to define it. The proposed model states that big data has three characteristics: Volume, Velocity and Variety. Volume indicates that a feature of big data is the huge amount of data produced and collected. Velocity indicates that the speed of data generation and collection should be very high. And the variability suggests that big data can be presented in a very diverse form and include semi-structured or unstructured data. (Laney, 2001) Thus, if the data has a huge volume measured in terabytes and petabytes, is collected in a short period of time or online, and includes a variety of data formats, such as audio, video, text information, etc., then such data can be considered big data.

In the future, the presented model received many developments, various researchers supplemented this model with new characteristics inherent in Big Data. Thus, in 2011, John Gantz and David Reinsel in their article proposed to consider the value of data as another characteristic of Big Data. (Gantz & Reinsel. 2011) In their definition Big Data is a “new generation of technologies and architectures, designed to economically extract **value** from very large **volumes** of a wide **variety** of data, by enabling high - **velocity** capture, discovery, and/or analysis". This model has become very widespread because it has been able to supplement the description of Big Data with a description of its necessity and purpose.

Attempts to describe and characterize Big Data did not stop there. Various authors have supplemented this model with new dimensions such as: Visualization, Veracity, Validity, Variability/Volatility and Complexity. (Patgiri & Ahmed, 2016) Such a variety of definitions may indicate that there is no consensus in the scientific community on the topic of definition and characteristics of Big Data. The many different approaches to defining big data show the importance of this area, but also speak to the challenges in this area.

The topic of challenges related to Big data is very much discussed and often raised in scientific papers. Ashabi, Sahibuddin, & Haghigh in their paper, reviewed the existing literature on current Big Data issues and challenges. (Ashabi, Sahibuddin, & Haghighi, 2020) As a result, they identified three categories of problems related to the topic of big data:

- Problems related directly to data (variety, volume, velocity etc.),
- Problems related to the process of working with data (how to capture data, how to change data, how to incorporate data etc.)
- Management problems (security, privacy, governance, ethical perspectives etc.)

According to the authors, further research is needed in the area of the described problems and the development of next-generation Big Data tools in order to fully take advantage of the benefits that Big Data provides. (Ashabi, Sahibuddin, & Haghighi, 2020)

Big data can really bring significant benefits and is used in a wide range of industries. Most often in scientific papers, among the industries with the greatest potential to benefit from big data, the following are distinguished: health, commerce, transport, tourism and political. In general, big data analysis provides an opportunity to obtain reliable information from the data, of course, if the data itself is representative. Such reliable information can be useful for decision makers to make the right decisions, and for companies to improve their policies and strategies, maximize profits and increase the competitiveness of enterprises. (Benjelloun, Lahcen, & Belfkih, 2015)

1.2. Data Analytics

Data analytics is a field of knowledge that consists of different approaches to collecting heterogeneous data from various sources, finding new knowledge from the collected data, and making predictions about possible events in order to innovate, gain competitive advantages, and make strategic decisions. (Davenport & Harris, 2007) However, the term data analytics is a general term that unites various approaches and methods of data analysis, such as online analytical processing (OLAP), data mining, visual analytics, big data analytics and so on. (Gudivada, 2017) In recent years, the term data science has become popular, but in fact it is the same field of knowledge as data analytics, only using more advanced data analysis methods such as neural networks or cognitive analytics.

Analytics can be different in terms of the data analysis methods used and the results of the analysis. One of the most popular frameworks that defines the types of data analytics is Gartner Analytics Ascendancy Model (GAAM). According to this model, data analytics can be divided into four types according to the difficulty and value of the analysis' results (Maoz, M. 2013). The four types of data analytics are as follows:

- Descriptive analytics: What happened?
- Diagnostic analytics: Why did it happen?
- Predictive Analytics: What will happen?
- Prescriptive Analytics: How can we make it happen?

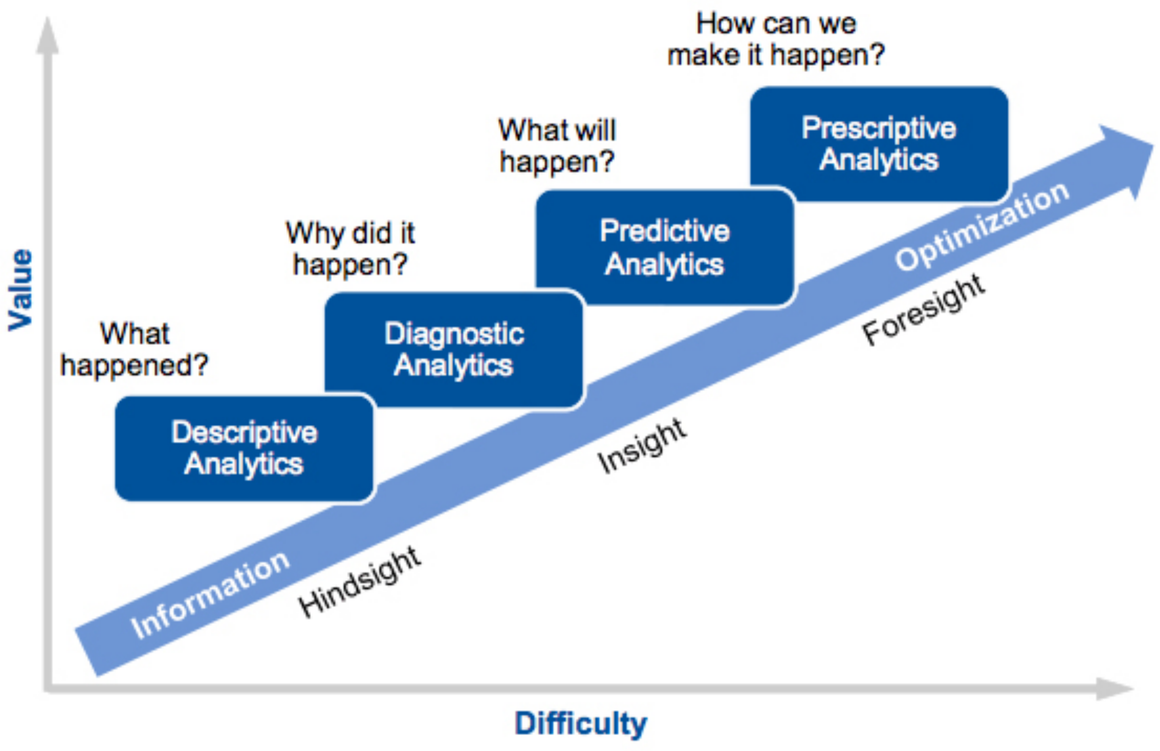

Figure 1. Gartner Analytics Ascendancy Model (Maoz, M. 2013)

Descriptive analytics answers the question "What happened?". Using descriptive analytics techniques to analyse data, you can understand what happened in the past. Examples of such analytics are the description of events that occurred using statistics, the construction of graphs and charts, and the preparation of reports. Dashboards and various data visualization systems are typical tools for such data analysis.

Diagnostic analytics is a more advanced form of analysis than descriptive analytics, which answers the question "Why did this happen"? This type of analytics allows you to take a deeper look at the data, comparing them with each other to establish a causal relationship. The most commonly used methods of diagnostic analytics are drill-down analysis, correlation analysis, pattern detection, and data mining.

Predictive analytics aims to answer the question "What will happen in the future?" based on the available data. This type of analytics can be called advanced analytics, as it requires the use of complex tools and methods of data analysis. One example of such analytics is demand forecasting, where a company uses historical data to predict what demand will be in the next month. If the predictive model is sufficiently accurate, the company can make decisions based on it. This type of analytics uses advanced tools, such as various programming languages such as R or Python, or special programs such as MATLAB or RapidMiner.

Prescriptive analytics is the most advanced level of data analytics. Prescriptive analytics answers the question "how do I make something happen?" Prescriptive analytics uses the results of predictive analytics, but in addition to this, feedback is used to study and improve the relationship between the actions taken and their results. Prescriptive analytics thus presupposes all favorable outcomes and suggests what courses of action should be taken to achieve a particular outcome. An example of such analytics can be various recommendation systems when a system based on user preference data offers movies, music, or products to other users with similar preferences.

This model implies that when using a more advanced type of analytics, the company becomes more analytically mature and can achieve more competitive advantages. (Maoz, M. 2013) But, the higher the level of analytics used, the more sophisticated tools are used to analyze the data. So, if for descriptive analytics companies can use simple data visualization tools, then for the prescriptive analytics, companies need to use advanced programming tools such as Python and additional packages. Therefore, to move to more advanced types of analytics, organizations must spend more resources on maintaining and developing analytical functions in the company. This leads to the fact that, compared to descriptive and predictive analytics, prescriptive analytics is even less mature in companies. (Gartner, 2019) This raises the question of how companies can approach the development of analytics more effectively in order to become more mature in analytics and start using more advanced types of analytics to achieve a greater competitive advantage.

1.3. Instruments and methods of data analytics

1.3.1. Data storage

Data warehouses are necessary in order to accumulate data from various sources and store data in a suitable form for further analysis. Once the data is accumulated in a data warehouse, it can be processed using various tools. There are different database architectures, the two main architectures are SQL and NoSQL. SQL is a structured query language that is used to manage data in traditional relational databases. In such databases, information is stored in the form of structured tables. This is an effective form of information storage in the case of small and well-structured amounts of information.

With the growing number of data, relational databases have become inefficient. So, there were non-relational databases, so-called NoSQL databases. (Cattell, 2011) Such databases can store unstructured data in the form of key - values, graphs, or documents. (McCreary & Kelly, 2013) NoSQL solutions are effective when working with huge amounts of data. The advantages of such databases are scalability and fault tolerance. In such databases, you can add and remove new data nodes at will without interrupting service. Also, such databases, due to the possibility of distributed data storage and processing, are safe from data loss, even if the hardware fails. (Carme & Jiménez, 2011).

With the growing interest in big data analytics, distributed computing systems such as Hadoop have become popular. This platform includes the Hadoop Distributed File System. In this file system, data is stored in blocks of equal size, and distributed across the nodes of a computing cluster (a group of computers). This type of data storage is effective for further data processing using other methods of Hadoop distributed computing. (Russom, 2013)

1.3.2. Distributed computing

With the growth of data volumes, the time for processing them using traditional approaches has increased significantly. The need for new approaches based on the principle of distributed computations has appeared . Working with data in parallel at the same time can provide high performance for intensive data operations (Russom, 2013). Such parrallel architecture distributes the work on data analysis between several blocks of processing, which reduces the process time.

A popular approach to distributed computing for working with data is the MapReduce model. MapReduce is a framework for computing certain sets of distributed tasks using a large number of computers (called "nodes") forming a cluster. The operation of this model consists of two stages: Map and Reduce. At the Map stage, the data is pre-processed by the master node (computer) of the system, this computer split the data and transmits it to other nodes in the cluster. At the next stage, each node separately processes the data and at the Reduce stage, the master node receives the result from the others and forms the final result, the solution to the problem. (Dean & Ghemawat, 2008)

Widely used tool for implementing the MapReduce principle is Hadoop. (Chen, Mao, & Liu, 2014) Hadoop is a set of utilities and frameworks for developing and executing distributed programs running on clusters of hundreds and thousands of nodes. Hadoop contains different packages for different tasks. For example, the Hadoop MapReduce package contains tools for programming distributed computing within the MapReduce paradigm. And HDFS (Hadoop Distributed File System) is a file system designed to store large files distributed among the nodes of a computing cluster. In such a file system, the distributed system is resistant to failures of individual nodes due to data replication. (Apache Hadoop 2021)

Distributed computing tools such as MapReduce or Hadoop are very efficient but at the same time easy to implement and use. (Chen, Mao, & Liu, 2014) A company needs a lot of resources to implement such systems in its data processing processes.

1.3.3. Tools for data analytics

Different data processing tools and are used at different levels of analytics. They differ both in cost and in the skills required to work with them. As we can see from Table 1, the most popular data analysis tools are Python, RapidMiner, R, and excel. (Piatetsky, 2019) Each of these tools are useful on different levels of analytics.

Software	2019 % share	2018 % share	2017 % share
Python	**65.8%**	65.6%	59.0%
RapidMiner	**51.2%**	52.7%	31.9%
R Language	**46.6%**	48.5%	56.6%
Excel	**34.8%**	39.1%	31.5%
Anaconda	**33.9%**	33.4%	24.3%
SQL Language	**32.8%**	39.6%	39.2%
Tensorflow	**31.7%**	29.9%	22.7%
Keras	**26.6%**	22.2%	10.7%
scikit-learn	**25.5%**	24.4%	21.9%
Tableau	**22.1%**	26.4%	21.8%
Apache Spark	**21.0%**	21.5%	25.5%

Table 1. Top Analytics/Data Science/ML Software in 2019 KDnuggets Poll (Piatetsky, 2019)

For descriptive and diagnostic analytics, relatively simple tools can be used. At these levels of analytics, Microsoft Excel can be a good tool for analyzing data. (Palocsay, Markham, & Markham, 2010) It allows you to build graphs and charts and perform simple data analysis. If a company needs more in-depth analysis, it can use special statistical programs, such as SPSS or STATA. These programs provide more tools for data analysis. Also, Different BI systems such as Tableau or QlikView gain popularity recently. These tools include dashboards, graphs, tables, and reports that are specially configured for convenient operational and strategic analysis and decision-making based on it. (Alade, 2017)

At the levels of predictive and prescriptive analytics, tools such as R, Python, MATLAB, or RapidMiner are used. (Bonthu & Hima Bindu, 2018) RapidMiner contains ready-made predictive models, which makes this tool more accessible to people without experience in advanced data analytics, while Python with its many libraries requires a huge amount of knowledge in mathematics and programming.

Python remains the most popular tool for big data analytics. (Piatetsky, 2019) Due to the large community of developers and a wide selection of libraries, this tool is very flexible and effective in building advanced models for data analysis. For example, at the levels of predictive and Prescriptive analytics, methods such as linear and logical regression, classification, clustering, decision trees, and so on are used. Python has special libraries and tools for working with these methods, for example scikit learn, matplotlib, NumPy, and so on. For more advanced data analysis methods such as deep learning, neural networks, python also has special tools, for example TensorFlow or Keras.

Thus, the implementation of simple analytics such as descriptive and diagnostic analytics requires much fewer financial costs, time and discipline from the company. Many companies can start using simple analytics in Excel or set up a BI system using data and resources that the company already has. With more advanced analytics such as predictive and prescriptive analytics, things are different. Such analytics require the involvement of a team of specialists with deep knowledge in mathematics, statistics, and programming. And building a full-fledged decision-making system based on data integrated into the company's existing IT systems, with systems for analyzing graphical and other information using neural networks, requires even greater financial costs. Such systems require a more responsible approach to data collection and storage. Data should be treated as a strategic asset. However, if companies are able to develop and implement such systems in various processes, they can gain significant advantages over competitors. (Henke et al., 2018)

1.3.4. Methods of Data Analytics

There are a huge number of methods for analyzing data. Often, the methods used for descriptive and diagnostic analysis are combined into one group. (Kaur & Phutela, 2018, Duan & Xiong, 2015) So we can distinguish three groups of methods for analyzing data by type of analytics:

- Descriptive and Diagnostic analytics methods
- Predictive analytics methods
- Prescriptive analytics methods

However, it is not possible to accurately divide all the methods into these three groups. Different methods can be used in several types of analytics at the same time. We can only separate the methods based on the frequency of use for a particular type of data analytics.

The simplest method of descriptive analysis is to study the correlation or trends of all possible attributes. For example, we can research popular products based on gender and age. These simple types of descriptive statistics are related to business intelligence and OLAP. (Duan & Xiong, 2015) In addition to these simple methods, there are more advanced methods that allow you to find hidden patterns in the data, such methods belong more to the category of diagnostic analytics. Kaur & Phutela's work identifies four such groups of methods for descriptive and diagnostic analytics. (Kaur & Phutela, 2018) These are:

- Associations
- Decision rules
- Cluster analysis
- Summarization Rules

In most cases, prescriptive analytics uses the results of predictive analytics. Therefore, it is difficult to divide the methods used in these types of analytics into two groups. (Lepenioti, Bousdekis, Apostolou, & Mentzas, 2020) In their paper, Lepenioti, Bousdekis, Apostolou, & Mentzas review the literature on predictive and prescriptive analytics. As a result, they define 25 methods used in predictive analytics and 38 methods used in prescriptive analytics. Not all of the methods described in this article are popular, many of them are used to solve very narrow problems. In another article, Duan & Xiong list the most popular methods used in predictive and prescriptive analytics. This list is as follows:

- Regression analysis
- Classification analysis
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Bayesian statistics
- Deep Learning algorithms

1.4. Organizational issues of big data analysis

Today, all organizations, regardless of whether they work in the industrial, public or non-governmental sector, understand the potential of data analysis. (Posavec & Krajnovic, 2016) But the introduction of data analytics brings not only advantages and benefits, but also causes a number of organizational problems.

Data analytics opens up great prospects for improving business processes in various organizational functions. In their article (Posavec & Krajnovic, 2016), identify the following business areas where data analytics can be applied:

- Marketing: decisions on pricing, locations of stores and branches, targeting promotions, web site customizations, digital media advertising placement;
- Finance: decisions on financial performance drivers, balanced scorecards, forecasts;
- Supply chain management: decisions on inventory, locations of distribution centers and warehouses, transport routing;
- Human resources: decisions on compensations, education, hiring employees;
- Research and development: decisions on product features, product effectiveness, product design;

In general, most authors agree on the scope of the data. The above business areas relate to the classic activities of the company directly or indirectly related to the product. In turn (Morabito, 2015) additionally highlights the organizational advantages that a company can achieve by implementing data analytics in its processes. According to them companies can:

- Improve decision making by lowering the cost of better-quality information analysis
- Improve business performance by disseminating information more effectively across the organization
- Improve collaboration by developing a common, enterprise-wide business intelligence, integrating views on identified business opportunities
- Generate and pre-test value propositions utilizing advanced and discovery analytics.

Thus, the scope of data analysis in organizations is very wide. However, the process of implementing data analysis is not easy and can lead to problems. Many authors have researched the challenges that companies face in implementing data analytics. (Morabito, 2015, Posavec & Krajnovic, 2016) Most often, among such problems, the following are distinguished:

- Lack of understanding of how to use analytics to improve business
- Lack of management bandwidth due to competing priorities
- Lack of analytical skills
- Culture that does not encourage sharing information
- Unclear ownership of data or ineffective data governance
- Lack of executive sponsorship

- Technology maturity levels

At the same time (Posavec & Krajnovic, 2016) emphasize the fact that data acquisition is not a problem right now. All the problems of companies are mostly managerial and organizational in nature.

The problem with employees is seen as the most significant for companies. (Morabito, 2015) Data analysts working with advanced analysis techniques are a rare resource. Such specialists should have a deep knowledge of mathematics, statistics and programming. Finding such expertise for companies is not easy and often very expensive, since such specialists have a high salary. Therefore, working with analytical competencies is important in the development of the company's analytical function.

Another frequently mentioned issue is the company's processes and structure. (Morabito, 2015) Advanced data analytics is an ever-evolving field, with new technologies emerging very frequently. In order to succeed in implementing new technologies in the company's analytics, you need to do it very quickly. Since in case of delay, a new technology may appear that will make all current developments irrelevant. Therefore, for a company, the transition to the use of advanced technologies in analytics is often associated with the adaptation of the organizational structure and processes to new circumstances.

Data analytics can be organized in several different ways in a company. It can be centralized, completely decentralized, or combine both options. In the latter case, analytics in the company, a centralized team of analysts can be organized in the analytics department, and data analytics specialists can also work in other departments. In the literature, there are disputes about the effectiveness of various options for organizing analytics. In a centralized team, analytics can be built in the most efficient and cost-effective way by centralizing resources. However, the disadvantage of a centralized team is the inability to create a culture of data-driven decision-making in the company, since analytics becomes the prerogative of a limited circle of people, and the other part of the company does not face analytics. (Saxena & Srinivasan, 2013) On the other hand, the literature mentions that the complete lack of centralization also negatively affects analytics. (Davenport et al., 2010) Therefore, in most cases, the best option for companies is a mixed analytics organization, when the company has a centralized analytics department and data analysis employees in other departments of the company. (LaValle et al., 2011)

In order for companies to effectively address the problems described above and improve their analytical function, they need to develop metrics for monitoring the current state of analytics.

(Cervone, 2016) Such a tool is the analytics maturity model, which allows companies to assess the current level of analytics development in the company and understand ways to improve it. (Morabito, 2015)

1.5. Analytics in Russian Companies

In 2018, the consulting company BCG estimated the volume of the big data market at 45 billion rubles. The average growth since 2015 has been 12 percent per year. (Ассоциация Больших Данных, 2018) According to the Russian Big Data Association, the big data market can reach 300 billion rubles by 2024. And indeed, more and more Russian companies are becoming involved in data analytics to one degree or another. In a study of 101 Russian companies with more than 500 employees, conducted by IDC with the support of Hitachi, more than 90% said that they face big data analysis to some extent. (Семеновская, 2019) Most companies that have embedded big data analytics said they plan to expand their dedicated analytics budgets. In general, from this study, we can conclude that Russian companies are actively involved in data analytics, see great potential in this area and plan to actively develop this competence in the company.

However, despite the great interest of companies in data analysis in Russia, the interest of the scientific community in this topic remains weak. There are no scientific papers devoted to the study of how Russian companies use data analytics. All research on this topic is published by companies or industry associations. Also, not a single scientific work was found devoted to the study of the maturity of Russian companies in the use of data analytics. However, this does not mean that such studies are not necessary. For example, a huge number of works on various aspects of the implementation and evaluation of analytics in companies have been published abroad.

1.6. Analytics maturity models

With the increasing use of data analytics in companies, the increasing amount of data collected, and the emergence of new tools and methods for data analysis, companies began to face challenges. As a response to these problems, many data analytics maturity models have been proposed. (Muller & Hart, 2016) Analytics maturity describes how deeply and effectively an organization uses tools, people, processes, and strategy to manage and analyze data to inform business decisions. Analytical maturity includes technology, data management, analytics, management, and organizational components. Maturity models are used to guide this transformation process. (Król & Zdonek, 2020)

In their work, Król & Zdonek (2020) conducted a literature study and identified the most well-known and popular models of analytical maturity. They looked at what criteria are used to assess analytical maturity and what levels of maturity are highlighted. A total of 11 analytical maturity models were analyzed:

1. Analytic Processes Maturity Model (APMM).
2. Analytics Maturity Quotient Framework.
3. Blast Analytics Maturity Assessment Framework.
4. DAMM - Data Analytics Maturity Model for Associations.
5. DELTA Plus Model.
6. Gartner's Maturity Model for Data and Analytics.
7. Logi Analytics Maturity Model.
8. Online Analytics Maturity Model
9. SAS Analytics Maturity Scorecard
10. TDWI Analytics Maturity Model
11. Web Analytics Maturity Model

As a result of the review of analytical maturity models, the Król & Zdonek (2020) come to several conclusions. On average, analytical maturity models distinguish 4 - 5 maturity levels. From companies that do not implement analytics to companies that are innovators in this field. Also, authors came to the conclusion is that most maturity models consider the following criteria that affect the level of analytical maturity:

- Data
- Analytics governance
- Goals
- Analysts
- Techniques

Similar conclusions are reached by Ariyarathna & Peter (2019) in their work. They identify two factors that affect analytical maturity, business related factors, and technical related factors. Among the business factors, they distinguish culture, performance management, business strategy, leadership and skills. Technical factors in their opinion include data management, integration, quality, governance and technology. Thus, it is important for companies to focus not only on the technological aspect of implementing analytics and data collection, but also to pay close attention to the organization of analytical work and competence management in the company. Table 2 summarizes the criteria that are important for the analytical maturity of the company, and provides a description of each of the criteria.

Criteria	Important issues
Data	Obtaining valuable and reliable analysis results requires that the company is able to collect high-quality data, and is able to integrate data from different sources and organize it. The data should be stored in such a way that it is quickly accessible and ready for analysis.
Analytics governance	It is important that the enterprise is focused on the management of analytics. This includes developing an analytics culture in the organization (the analytics ecosystem), developing and implementing an analytics strategy, and adopting analytics goals. In analytical organizations, there are leaders who take full advantage of analytics and guide the development of the organization in such a way that it uses the potential of data analysis. Therefore, it is important that the analytics implementation initiative comes from the top level. This increases the adoption of an analytics culture across the enterprise and simplifies the implementation of analytics initiatives.
Goals	Analytics should be adapted to the company's strategic and corporate goals. To do this, the company must make appropriate plans, both long-term and short-term.
Analysts	The organization must have a broad set of data analyst competencies that can work with both simple tools and advanced data analysis techniques. To do this, the company needs to effectively manage data competencies, try to attract world-class specialists, and invest in additional employee education.
Techniques	The company's ability to use a wide range of technologies in the field of data analytics is important. This includes various data analytics tools, such as spreadsheets, programming languages, or non-specialized software packages. It is also important what data processing methods companies use, it is important that the company is able to take advantage of existing data analysis methods.

Table 2. Important issues for data analytics maturity criteria. (Davenport & Harris, 2017; Halper & Stodder, 2014)

However, despite the variety of existing models of analytical maturity, many authors emphasize the need for further research in this area. Król & Zdonek (2020) argue that further development of new models of analytical maturity is needed in the future. In particular, specially adapted for various sectors and industries. Muller & Hart (2016) also argue that with the advent of new methods and technologies for data analysis, models of analytical maturity should be adapted to new technologies.

Several authors raise the question of the need to adapt the models of analytical maturity to the conditions in which the company operates. Thus, Nda, Abdul Hamid, & Tasmin (2020) argue that analytical maturity can be influenced by economic adjustment, cultural differences, and other challenges. An analytical maturity model should take these factors into account. Ariyarathna & Peter (2019) in their work say that most of the existing models of analytical maturity were developed in the conditions of developed countries. Such models may not be effective when applied in emerging markets. As mentioned earlier, not a single research article or work was found on the study of the problem of maturity of Russian companies in data analytics, or the development of an analytical maturity model for Russian companies. Therefore, it is necessary to continue research in the field of analytical maturity models, and to develop models that take into account the characteristics of emerging markets.

1.7. Research gaps and research questions

From the literature reviewed on the topic of data analytics, we can draw several conclusions. First of all, technologies in the field of data analytics are actively developing, companies can collect more and more data and benefit from their analysis. But the introduction of data analytics, especially advanced ones such as predictive and prescriptive, causes many problems for companies. Companies need to take a more careful approach to hiring employees and organizing processes in the company for effective implementation of analytics. To do this, companies need to understand their weaknesses and understand what needs to be done to be more advanced in data analytics.

For these purposes, there are models of analytical maturity. These tools can help companies analyze their current situation with analytics and identify weaknesses that require attention. Using this tool, companies can become more mature in the field of analytics. There are many such models, but not all of these models can be called effective for Russian companies with a high degree of probability, since they were developed in conditions of other markets. There is no model of analytical maturity developed specifically for the conditions of the Russian emerging market.

And it is important that such a model is adapted to the conditions of the economy, political and cultural characteristics of the market in which the company operates.

Therefore, it is necessary to conduct research in the field of analytical maturity of Russian companies. It is necessary to investigate the possibility of applying existing models of analytical maturity in Russian companies. It is necessary to draw conclusions about whether Russian companies using data analytics have any significant differences that prevent the use of existing models of analytical maturity. It is also necessary to develop an analytical maturity model adapted specifically for Russian companies, but the process of developing an analytical model is a very complex and lengthy process. (Bruin et al., 2005) Therefore, before starting to develop such a model, it is necessary to carefully study the current situation and maturity criteria of Russian companies in the field of analytics. The purpose of the study and the research questions can be formulated as follows based on the identified research gaps.

The research goal of the thesis is to understand whether the existing models of analytical maturity are suitable for use in Russian enterprises, and to investigate the current situation with the analytical maturity of Russian enterprises.

The **research questions** can be formulated as follows:

RQ1 How do Russian companies currently use data analytics?

RQ2 Can the existing models of analytical maturity be considered applicable for Russian companies?

RQ3 What levels of analytics maturity can be determined for Russian companies?

RQ4 What are the characteristics of companies at each maturity level?

RQ5 What Russian companies need to do to get to the next level of maturity?

CHAPTER 2. METHODOLOGY

This section is dedicated to developing the design of the study to answer the formulated research questions. At the beginning, we will describe the general approach to the study and what stages it will consist of. Next, each stage will be described in detail, including the justification of research method, the formulation of hypotheses, and justification of approaches and methods of results analysis.

2.1. Research Methodology

To answer the research questions, a qualitative study will be conducted in the form of a structured survey. The study will be conducted in five stages, graphically shown in Figure 2. At the first stage, the literature on data analytics and the analytical maturity of the company was reviewed. Five criteria affecting the analytical maturity of companies was identified, and these criteria was described. Further, based on the identified gaps in existing research on the company's maturity in the field of data analytics, research questions, that require an answer as a result of this study, were compiled.

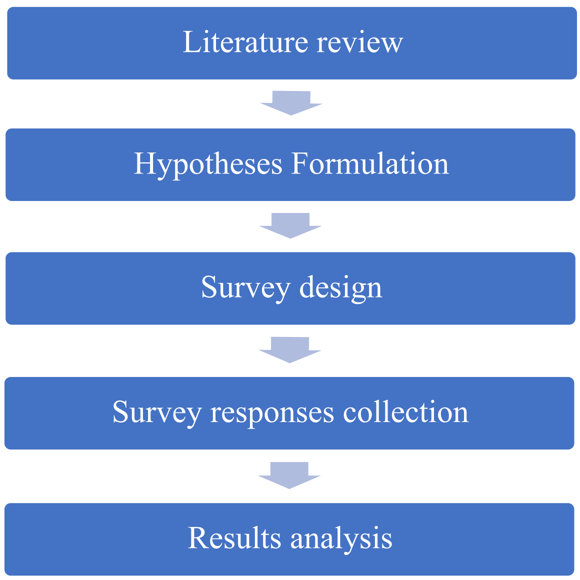

Figure 2. Research process

At the second stage, research hypotheses will be formulated in order to answer the first two research questions. These hypotheses should be formulated using literature that provides a theoretical explanation of the relationship between different variables in the field of data analytics. If these hypotheses are confirmed, it will mean that Russian companies use data analytics in accordance with the existing theoretical concepts of how companies use data analytics. Since the existing literature on how companies use data analytics is mainly based on the research of foreign companies, the confirmation of the hypothesis will indicate that the use of data analytics in Russian companies in general does not differ from how foreign companies use data analytics. Accordingly, we can conclude that the existing models of analytical maturity can be used in Russian companies. If the hypotheses are not confirmed, it will indicate that Russian companies have some peculiarities in using data analytics.

At the third stage, a survey with fixed answer options will be developed. The survey will be structured taking into account the identified five criteria that affect the company's maturity in analytics, and taking into account the compiled research hypotheses. Such a survey will allow to test the hypotheses formulated, and describe the features of Russian companies in the use of analytics, and how Russian companies use data analytics.

In the fourth stage, the responses to the survey compiled in the previous stage will be collected. To do this, the target audience of the survey will be determined, and the size of the required sample will be determined. The channels and methods for collecting responses to the survey will be defined. The results of the survey collection will also be summarized including the total number of survey requests sent out and the response rate.

At the fifth stage, the analysis of the results will be carried out to answer the questions of the study. The analysis of the results will be carried out in two stages. At the first stage, the first two research questions will be answered, and the characteristics of data analytics in Russian companies will be described. There will also be a test of research hypotheses, in order to find out whether there are any features of using data analytics in Russian companies. In the second stage of the results analysis, all companies will be divided into groups with similar characteristics of using data analytics, and as a result, being at the same level of analytical maturity. A description of the company in each group will be given. There will also be recommendations for companies in each group. The recommendations will contain practical tips on how to move to the next group of analytical maturity, that is, to become more advanced in the use of data analytics.

2.2. Conceptual research model

To answer the research questions, it is necessary to investigate how Russian enterprises use data analytics, according to the criteria that affect the analytical maturity of the enterprise. As discussed in the first chapter, there are many models of analytical maturity that help to determine maturity level of a company in data analytics. However, for the purposes of our research, it seems impossible to take the existing model of analytical maturity for the study of Russian companies, since it was found that there is a possibility that the existing models of analytical maturity are not suitable for use in Russian companies. Therefore, a new model was developed for the purposes of this study. As a result of a review of the literature, it was found that in most existing models of analytical maturity, the main criteria for analytical maturity are data, analytics governance, analytical goals, analysts, and technologies. Therefore, in this study, studies of how Russian enterprises use analytics will be made taking into account these five criteria. So, the study model is provided in Figure 3. This model also reflects the hypotheses set for each criterion of analytical maturity. Using this model will allow us to comprehensively investigate how Russian enterprises use data analytics, and answer the questions raised by the study.

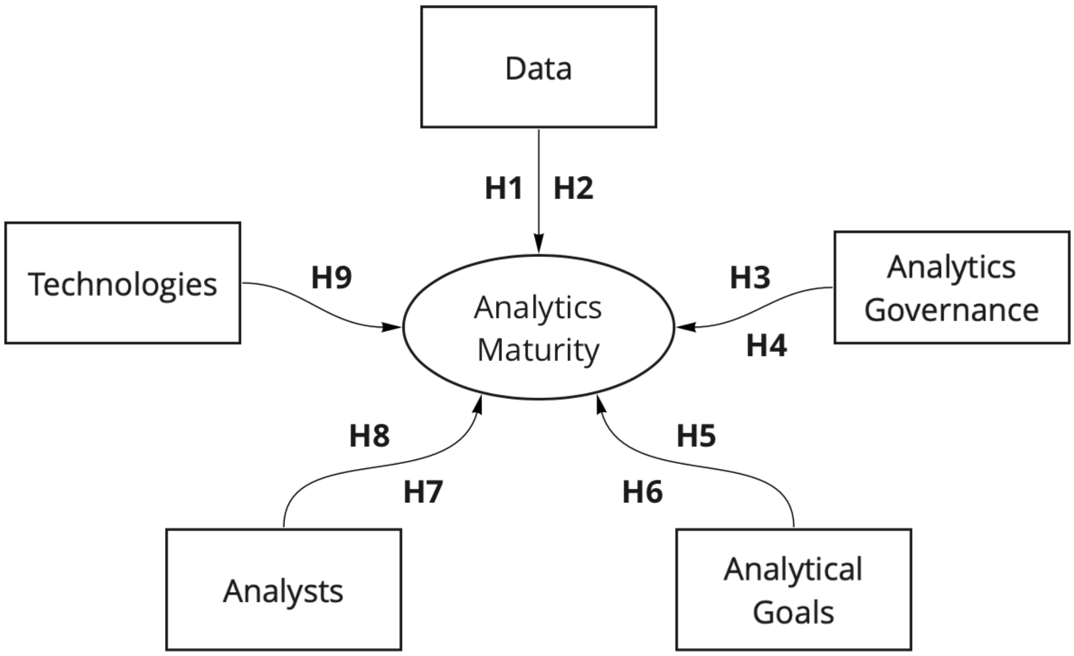

Figure 3. Research model

Further in this section, the justification of the chosen research methodology will be presented. The methods and methods of empirical research will be considered and the reason for the choice of the survey will be given. The methods of analyzing the survey results will also be considered, and the reason for choosing regression analysis and cluster analysis will be given.

2.2.1. Data collecting method justification

To answer the research questions, it is necessary to collect data. The approach for data collection depends on what data should be collected. In our case, the data will be primary, since as was previously mentioned in the literature review, there is not much research on the use of data analytics in Russian companies. Therefore, we cannot use the information already available to conduct the research. The selected data should be descriptive and not experimental, as one of the objectives of the study is to describe how Russian companies use data analytics. Categorical data were selected for the purposes of this study, as this type of data allows us to use data collection methods which easier for respondents to participate in. Thus, we will be able to collect more responses. The study will contain categorical variables, both nominal and ordinal, which will allow using statistical methods of data processing to answer the research questions.

Thus, having determined the types of data that we are going to collect, we can decide on the method of research. In total, there are six most common methods of research:

- Experiment
- Interview
- Case study
- Survey
- Literature review

However, not all of these methods are suitable for the purpose of our study. Literature analysis is also not suitable for us, since the type of data collected in this method of analysis is secondary data, but we need primary data. The experiment is not suitable for us because we need descriptive data, and this type of research allows us to collect experimental data, which will not help to solve the goal of our research, to describe how analytics is used in Russian companies at the moment. The interview is suitable for us, but using this type of data collection, we will only be able to collect a few opinions, thus not covering a wide range of Russian companies, so this method is not suitable for us. Case studies allow us to collect very deep information about a small number of research objects. In our case, we need to study Russian companies with different levels of data analytics usage, which is almost impossible to do with a case study. Thus, the most optimal method of research in our case is a survey. The survey will allow us to explore the opinion of employees working in Russian companies, about how these companies work in the field of data analytics.

2.2.2. Research hypotheses formulation

To answer the first two questions of the study, it is necessary to formulate research hypotheses. Research hypotheses were formulated for each of the five criteria that affect a company's maturity in using data analytics. These hypotheses were formulated using information about each of the criteria, provided in Table 2, and the information given in the literature describing how different aspects of the implementation of data analytics in the company affect each other. If these hypotheses are confirmed, it will mean that Russian companies use data analytics in accordance with the existing theoretical concepts of how companies use data analytics. Since the existing literature on how companies use data analytics is mainly based on the research of foreign companies, the confirmation of the hypothesis will indicate that the use of data analytics in Russian companies in general does not differ from how foreign companies use data analytics. Accordingly, we can conclude that the existing models of analytical maturity can be used in Russian companies. If the hypotheses are not confirmed, it will indicate that Russian companies have some peculiarities in using data analytics.

Hypotheses about the data

How much data a company is able to collect, and how different the types of data collected are, is a direct indicator that affects the maturity of a data analytics company. (Davenport, T. H., & Harris, J. G., 2017) Therefore, it is important for us to understand how Russian companies do this. Lismont, Vanthienen, Baesens, & Lemahieu (2017) found that the wider the use of company data analytics, the more types of data companies collect. Therefore, the first hypothesis about the data can be formulated as follows.

H1: The more a company uses analytics, the more types of data it collects

Another important aspect that affects how many types of data a company collects is how the company's data collection processes are organized. Companies that have standardized the process of collecting and processing data, and in particular plan in advance what data should be collected, usually collect more different types of data. (Davenport, T. H., Harris, J. G., & Morison, R., 2010). So, the second hypothesis from the data can be formulated as follows:

H2: The number of types of data collected depends on whether the company determines in advance what data should be collected

Hypotheses about the analytics governance

An important aspect is how companies organize data analytics. In more detail, the different types of organization of data analytics were discussed in section 1.4. Many aspects can affect how a company organizes data analytics, one of these aspects is how widely the company applies data analytics. (Davenport, T. H., Harris, J. G., & Morison, R., 2010) Thus the first hypothesis about the organization of data analytics can be formulated as follows:

H3: How analytics is organized in a company depends on how widely analytics is used

Another important aspect that affects how a company organizes data analytics are how many employees are engaged in data analysis. Sometimes, if a company employs a small number of employees in data analysis, it doesn't make sense for the company to use a mixed type of data analytics organization. (LaValle et al., 2011) Thus, the second hypothesis about the organization of data analytics can be formulated as follows:

H4: How analytics is organized in a company depends on how many data analysts work in the company.

Hypotheses about the goals in analytics

An important aspect of the implementation of data analytics is the planning of analytical activities. If the company has short-term and long-term data analysis planning, then the company can get more benefits from implementing data analysis. (Davenport, T. H., & Harris, J. G., 2017) Therefore, the first hypothesis about the company's goals in analytics can be formulated as follows:

H5: How much benefit a company gets from implementing data analytics depends on how it plans data analytics

Also, how much benefit a company gets from implementing data analytics is influenced by how the company approached planning the implementation of data analytics prior to its implementation. In the literature, it is mentioned that the more responsible a company approaches the process of planning the implementation of data analytics, the more benefits it can get. (Davenport, T. H., & Harris, J. G., 2017) Thus, the second hypothesis about the company's goals in analytics can be formulated as follows:

H6: How much advantage a company will gain from implementing data analytics depends on how many goals the company had before implementing data analytics

Hypotheses about the analysts

How effectively a company manages the competencies of data analysts is an important aspect in the application of data analytics in companies. With the growing size of analytical departments and tasks in data analysis, the importance of how the company interacts with data analysts, and work with competencies, is growing. In the literature, it is mentioned that the more effectively a company manages the competencies of analysts, the more effective and wider the use of data analytics. (Halper, F., & Stodder, D., 2014) Thus, the first hypothesis about analysts can be formulated as follows:

H7: How effectively a company manages data analytics competencies depends on how widely the company applies data analytics

Another important aspect is how many employees work in the company in the field of data analytics. When analytical departments become large, this imposes additional obligations on the company in the field of managing employee competencies. Accordingly, in such companies, the approach to managing the competencies of data critics is more advanced and responsible. (Davenport, T. H., & Harris, J. G., 2017) Thus, the second hypothesis about data analysts can be formulated as follows:

H8: How effectively a company manages the competencies of data analysts depends on how many data analysts work in the company

Hypotheses about the analytics techniques

From the point of view of technology, what methods companies use to analyze data is important. The number of such methods can serve as an indicator of the company's analytical maturity. If companies use too few methods of data analysis, this may indicate that the company uses either only advanced or only simple methods of data analysis, which is not the most effective way to use data analytics. (LaValle et al., 2011) In the literature mentions that the more companies use data analytics, the more data analysis methods are used. (Davenport, T. H., & Harris, J. G., 2017) Thus, the first hypothesis about data analysis techniques can be formulated as follows:

H9: The number of data analysis methods used depends on the width of the analytics usage

2.2.3. Justification of the hypothesis testing method

In order to test the hypotheses, put forward, it is necessary to study the correlation between the variables. Each of hypotheses formulated based on the assumption of a relationship between two variables. If we can prove that there is a statistically significant relationship between two variables in one of the hypotheses, then we will assume that the hypothesis is proven. If such a statistically significant relationship is not found, then we will assume that the hypothesis can be rejected. In order to study the relationship between variables, we can use two methods:

- Linear regression
- Correlation analysis.

Both methods of analysis are suitable for our purposes. Both linear regression and correlation analysis allow us to study the direction and extent of the relationship between the two variables. However, regression analysis provides more options. Regression analysis, in addition to finding all the independent variables that affect the dependent variable, also allows us to predict the value of the dependent variable in the future. For our purposes we will use stepwise multiple linear regression approach, as it allows us to select independent variables in several stages, removing at each stage all independent variables that are not statistically significant.

There are two types of stepwise multiple linear regression, forward, backward. Forward variable selection starts without candidate variables in the model. At the beginning, the variable with the largest R squared is selected. At each subsequent step, the variable that most increases R squared is selected. Variables stop being added if none of the remaining variables is significant. Once a variable enters the model, it cannot be deleted. The backward selection model starts with all the candidate variables in the model. At each step, the variable that is least significant is deleted. This process continues until there are no non-essential variables left. We can set the significance level at which variables should be removed from the model. For our research purposes, the backward stepwise multiple linear regression approach is suitable. As a result of multiple linear regression, only the independent variables that correlate with the dependent variable remain at the required level of statistical significance, in our case < 0.05 (p-value). For the purposes of this analysis, the SPSS Statistics program will be used.

2.2.4. Justification of the method for determining the levels of analytical maturity

To answer the remaining questions of the study, it is necessary to determine the levels of implementation of data analytics of Russian companies. To do this, it is necessary to solve the clustering problem, as a result of which Russian companies will be divided into several groups, each of which will contain companies with similar characteristics. To do this, we need to determine which approach should be used for clustering, and determine which variables will be used for clustering.

For clustering purposes, variables were selected that reflect the breadth of the company's data analytics application, the way the company's data analytics is organized, and the size of the analytical departments. In addition, variables were used to reflect what data analytics methods companies use. This choice of variables is due to the fact that we need to divide companies by the levels of implementation of data analytics in the company. And the breadth of data analytics implementation and the data analytics methods used directly reflect the level of data analytics implementation in companies. By dividing the companies according to these criteria, we will be able to study the characteristics of the companies in each group according to the other criteria of the company's maturity in data analytics.

When we have defined the changes that will be used for clustering, we can proceed to the choice of the clustering method. Cluster analysis or clustering is the task of grouping a set of objects so that objects in one group are more similar to each other than objects in other groups. There are a huge number of approaches to solving clustering problems and many different methods. There are four popular models of clustering, each of which contains different clustering algorithms.

- Hierarchical clustering
    - Agglomerative hierarchical clustering
- Centroid-based clustering
    - K-Means
    - PAM (K-medoids)
- Distribution-based clustering
    - Expectation maximization clustering
- Density-based clustering
    - DBSCAN

To solve our goals in each clustering model, we have identified several of the most popular algorithms. However, not all algorithms are suitable for our purposes, since not all algorithms can determine in advance the number of clusters into which the objects under study should be divided. Therefore, only three algorithms were selected for further tests, K-Means, PAM (K-medoids), and Expectation maximization clustering. For each of the selected algorithms, tests were conducted to find out which algorithm is more suitable for our purposes. Clustering of the collected data was performed using each of the algorithms. And in the future, the Chi Square test was used for the results obtained, in order to find out how effectively the companies were divided into clusters. This test allows us to evaluate how objects companies in one group differ from companies allocated in another group by clustering variables.

As we can see from the results of these tests shown in Table 3, the most effective method was PAM (K-medoids). The results of clustering using this method, there is a significant difference (p-value < 0.05) for each of the variables between the obtained groups of companies. Other algorithms were not so effective. The results of the Chi Square test specifically for this clustering algorithm can be found in Appendix B. Therefore, for the purposes of our research, we will use the PAM (K-medoids) method.

Clustering method	**Number of p-values > 0.05 (Chi Squared test)**
K-Means	2 / 23
PAM (K-medoids)	0 / 23
Expectation maximization clustering	8 / 23

Table 3. Test results of clustering methods

2.3. Expected findings

As a result of this study, two main results are expected. The first result is that it will be found out whether the existing models of analytical maturity can be applied in Russian companies. If all the hypotheses are confirmed, it can be concluded that Russian companies do not have serious differences from foreign companies in the use of data analytics, and Russian companies can use existing models of analytical maturity. The second result of the study will be to determine the level of maturity of Russian companies based on the opinions of employees working in these companies. For each maturity level, the characteristics of the companies included in this level will be identified, and the data analysis challenges faced by the companies at each level will be identified. At the end, there will be recommendations for companies at each maturity level on what steps should be taken to become more mature in the field of data analysis. This work will serve as a basis for further research of Russian companies and aspects of the implementation of data analysis in them. This work will contribute to the further development of maturity models for Russian companies.

CHAPTER 3. EMPIRICAL STUDY

3.1. Survey design

To answer research questions and test hypotheses, a survey was compiled. In order to reduce the time required to complete the survey and obtain data in a format suitable for further statistical research, the survey contained mostly multiple-choice questions. The survey aims to examine the opinion of employees working in Russian companies about how the company applies data analytics. The Google Forms platform was chosen for the survey, as it is cross-platform and the most convenient option for most people.

The survey was compiled taking into account five factors that affect the company's maturity in data analytics identified earlier. This survey structure allows us to consider in detail the respondents opinion on all the characteristics of the company that affect the maturity in analytics. Since most of the existing models of analytical maturity are commercial products, it was not possible to borrow questions from one of the models. Therefore, the survey was compiled using the description of the criteria that affect the maturity of data analytics provided in Table 2. The sources of this description are Delta Plus Model (Davenport & Harris, 2017) and the TDWI Analytics Maturity Model (Halper & Stodder, 2014). Finally, survey contained 50 questions structured in seven sections:

- Questions about company
- Questions about Data
- Questions about Analytics governance
- Questions about Goals
- Questions about Analysts
- Questions about Technologies
- Questions about the survey participant

The first question block contains questions about the company, such as the industry, the approximate size of the company by employees, and revenue. This block of questions is intended to give a rough understanding of which companies participated in the survey. The name of the company was also collected during the survey, but the results of the survey do not contain conclusions about the activities of a particular company, and this information is not used in the study in any way.

The second set of questions is intended to provide an understanding of how in the respondent’s opinion the company handles data. This part of the survey contains a question about how the company collects data what types of data it collects and how the process of working with data in the company is managed. The third set of questions contains questions about how data analytics is managed in the company and should give an understanding of how the company's data analytics management process works. This section contains questions about how company data analytics is organized and funded, who manages data analytics, and how widely data analytics is used in the company. The fourth set of questions contains questions about the company's goals of analytics, and is intended to give an understanding of how the company plans its activities in the field of data analytics. This block of questions contains questions about how the company plans its analytical activities, if the company has long-term or short-term planning, and contains a block of questions about the planned results and the actual results of implementing data analytics. The fifth set of questions contains questions about how the company works with data analysts. This set of questions is intended to answer the question of how effectively the company manages the competencies of data analysts, and contains questions about how the company is involved in the process of additional education of employees engaged in data analytics, and whether the company has competencies in various areas of data analysis. The sixth block contains questions about the technologies used in data analytics. This set of questions should give an idea of what methods and tools companies use for data analytics and also contains questions about what areas data analytics takes and what types of data analytics are used.

The last seventh question block contains questions about the survey participants. The block of questions should give an understanding of who took part in the survey and what positions these employees occupy. Also, all participants of the survey were given the opportunity to express their opinion about the survey, and leave their contacts if they are interested in receiving the results of this study.

After the survey was compiled, the final version was sent to three data analysis specialists working in large Russian companies as senior data analyst and project manager in the field of data science, to confirm that the survey was compiled correctly and can be used in further research. During the discussion, some adjustments were made to some of the answer options and the questions themselves, in particular, several methods of data analysis were added and removed. Communication with them took place in an unstructured form, mainly in personal messages. In general, they confirmed that the survey was compiled correctly, and correctly reflects the aspects of the use of data analytics in Russian companies.

A full list of questions used in survey is provided in appendix B.

3.2. Data collection

Since the purpose of the survey is to study the opinion of employees working in Russian companies about how the company applies data analytics, the target audience for this question is employees engaged in data analytics, since the opinion can adequately reflect how the company works in the field of analytics. Therefore, the target audience is the top management (heads of departments and divisions, preferably data analysis departments) of large Russian companies, as well as employees of data analysis departments, data analysts and project managers in the field of data analysis. The opinion of this audience will allow to collect objective data on various aspects of the company's maturity in the field of data analysis.

To determine the size of the required sample for the survey, it was assumed that the target audience of the survey is large and medium Russian enterprises. According to the service of statistical information on Russian legal entities, in 2020 there were 30,000 large and medium-sized enterprises in Russia. (СПАРК – Проверка, анализ и мониторинг компаний 2021) These companies were chosen for the survey because they have started the process of forming analytical departments and these companies are no longer at the zero level of using data analytics. Using the formula to calculate the required sample, taking into account the 95% confidence level and 10% margin of error, it was found that the required sample size is 80.

A search was conducted for potential platforms on the Internet that can provide access to the necessary audience for this issue. All the most popular social networks, such as VK, Facebook, Odnoklassniki and so on, were considered. The relevant telegram channels and communities of data scientists were also investigated. It was found that the largest number of potential respondents are available in the social network VK. The survey was conducted over a period of one month. A total of 735 survey requests were sent out. As a result, 80 responses were collected. The response rate was 10.88%.

3.3. Description of companies and survey participants

In total, 80 Russian companies of various sizes and operating in different industries took part in the study. Figure 4 shows information about the industries in which the companies operate. Since many companies work in several industries at once, this figure shows how many companies have indicated that they work in a particular industry. From Figure 4, we can see that the most frequent industries are Information Technology, Financial Services, Retail and Online Retail.

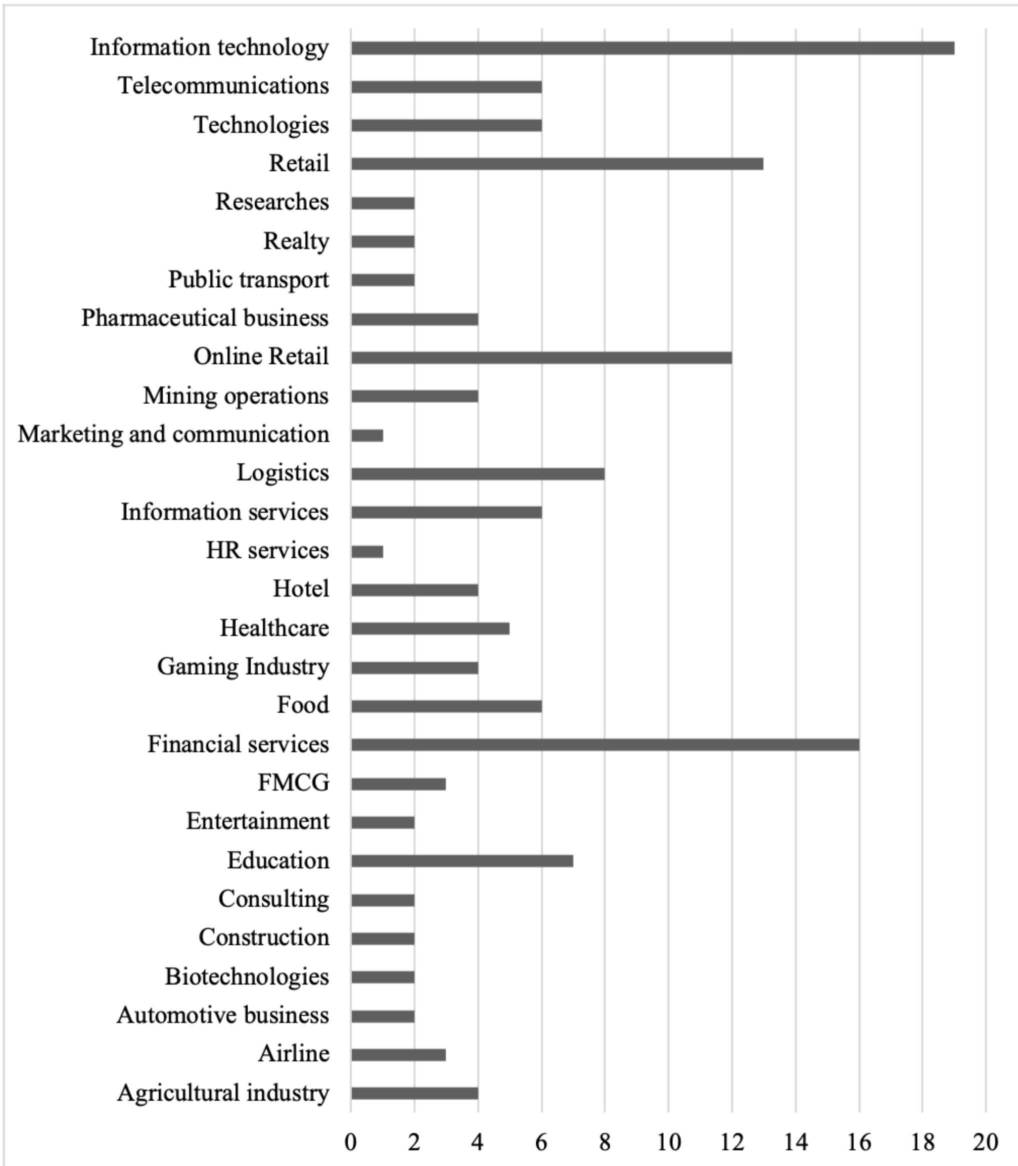

Figure 4. Industries of the companies in which the respondents worked

The information obtained by industry can indicate which industries are developing the most data analytics. So, companies working in the field of retail and online retail are often very large companies, with many processes, which is the reason that business analytics is developed in them. Also, companies engaged in the field of finance and banking have developed analytics, as in most cases they have developed online services. And companies are engaged in the field of information technology by virtue of their field in which they work are involved in data analytics.

Also, companies vary greatly in the number of employees working in them. From Figure 5, we can see that a large share of companies employs between 1,000 and 10,000 people. We can also see that the survey was attended by companies that employ more than 100,000 people, one of the companies employs about 700,000 people.

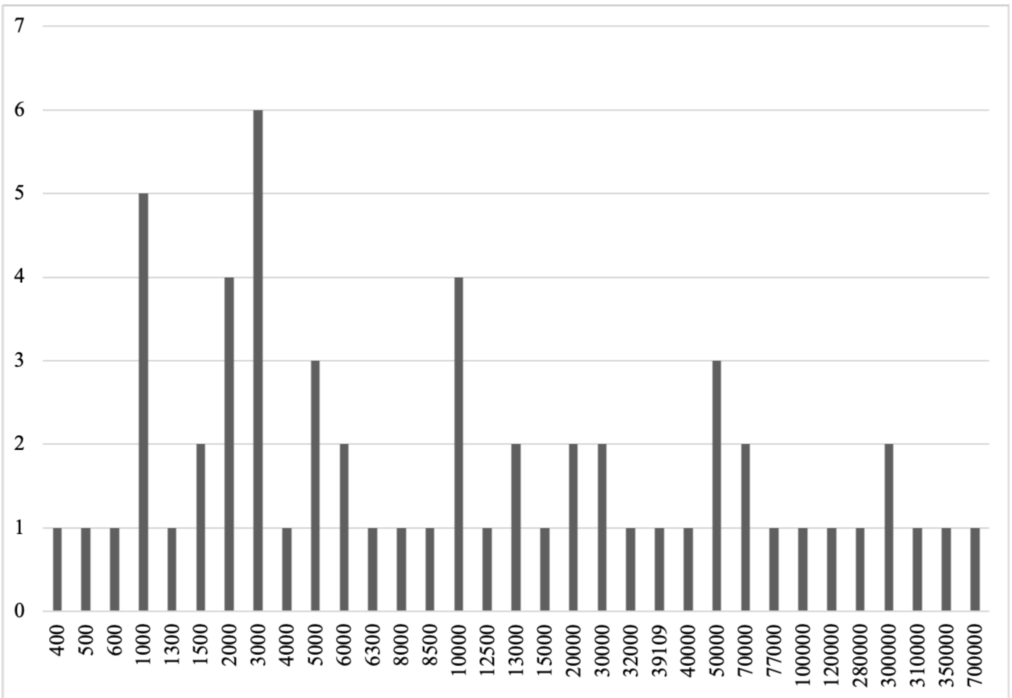

Figure 5. Companies by number of employees

Information about the employees who took part in the survey, and in particular what positions they occupy, is provided in Figure 6. In total, three categories of company employees took part in the survey. The largest category of employees were data analysts and data scientists, with 67 participants. The second largest category was senior managers of companies, there were 9 such people, and 4 heads of the analytical department took part in the survey.

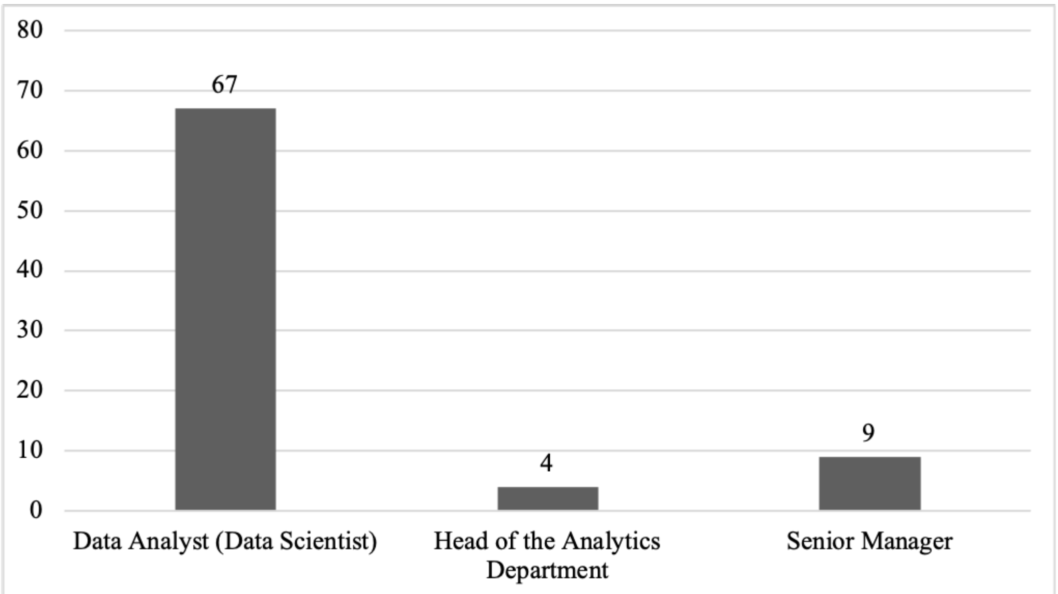

Figure 6. Positions of employees who participated in the survey

The full list of companies participating in the survey can be found in Appendix C.

CHAPTER 4. RESULTS ANALYSIS

4.1. Results of descriptive analysis and hypotheses testing

4.1.1. Data perspective

According to Kiron, Shockley, Kruschwitz, Finch, & Haydock (2012), an important component of the work of companies with analytics is the ability to accumulate data from various sources. Advanced companies in analytics collect data not only from internal sources but also look for data outside the company. According to the survey results, 81.5% of the participants said that their company uses all possible data sources both inside and outside the company. 13.5% of the participants said that the data is collected only within the company. Such results may indicate that companies are trying to get as much data as possible for further use. These conclusions are confirmed by the results of the survey. 73.8% of respondents said that their organization strives to collect as much data as possible. 76.3% of respondents said that their company is constantly looking for new data sources. 83.8% of respondents said that their company treats data as a strategic asset, while only 67.5% admitted that every employee understands the value of data for business. This may indicate that the company is taking the data collection process seriously. This is confirmed by the fact that 85% of respondents said that their company has clear procedures for collecting and storing data.

From the analysis of the survey results, we can see that the most popular type of data that companies collect is the purchase history, 83 % of the companies collect this data. Another popular type of data is data about the company's internal processes, such as conversion or delivery times, or customer feedback, such as surveys. These types of data are collected by about 75 % of companies. Only about 40% of companies collect data such as images, audio recordings, and videos. The most unpopular type of data is environmental data such as weather. In general, from the results obtained, we can conclude that the companies are trying to collect common data generated within the company, which is stored in the usual formats, and which is clear how to analyze. Unstructured data such as image, video, and audio recordings require more advanced analytics techniques, so they are collected with fewer companies.

To test the hypotheses about the data, formulated earlier, a backward stepwise regression analysis was performed. The dependent variable was the number of data types used in the company. All other variables except the data types were selected as independent variables. In total, two previously formulated hypotheses about the company's analytics goals were tested:

H1: The more a company uses analytics, the more types of data it collects

H2: The number of types of data collected depends on whether the company determines in advance what data should be collected

The results of the regression analysis are shown in Table 4. We can see the dependent variable depends on six independent variables with a significance level less than 0.05. We can see that the number of data types collected depends on the width of the use of data analytics with a coefficient equal to -1.633. The coefficient value is negative because the answer options were encoded in descending order, that is, zero is the use of data analytics by all departments of companies, and two is the use of analytics only for a few initiatives or projects. Thus, we can confirm the first hypothesis.

Also from Table 4, we can see that the number of data types used depends on whether the company determines in advance what data should be collected, with a coefficient of -3.151. The coefficient is negative since the answer options were encoded in descending order, that is, zero is if the company determines in advance what data yes should have been collected and one if not. Thus, the second hypothesis is also confirmed.

We can also see that in addition to the variables defined in the formulated hypotheses, the number of data types collected depends on the other 3 variables with a significance level less than 0.05. Thus, the number of data types collected also depends on whether the company evaluates and regulates the quality of the collected data, and if the company has regulations describing the process of data collection and analysis. The number of data types collected is also affected by the number of data analysis tools used, with a coefficient of 1.535. Which tells us that the more tools a company uses, the more types of data it collects.

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
.846	.715	.677	2.152

Coefficients[a]

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	5.130	1.030		4.983	.000
The company evaluates and regulates the quality of data	2.715	.796	.259	3.409	.001
The organization determines in advance what data should be collected	-3.151	.760	-.325	-4.147	.000
The company has regulations and rules that describe the process of data collection and analysis at all stages?	-2.729	.646	-.336	-4.224	.000
sum_of_instruments	1.535	.211	.574	7.284	.000
How widely is data analytics used in your company?	-1.633	.389	-.319	-4.195	.000

a. Dependent Variable: sum_of_data_types

Table 4. Regression results for data issues

4.1.2. Data analytics governance perspective

From the point of view of the organization of analytics in companies, it is important to know how widely the company uses analytics. So, the result of the survey analysis shows that 45 % of the company uses analytics in all departments of the company and in many processes. 36 % of companies use analytics only in a few departments, 19 % of companies use analytics only in a few initiatives and projects.

As described in section 1.4, there are different ways to organize analytics, the centralized way, the decentralized way, and the mixed form of analytics. From the analysis of the result, we can see that the centralized form of organization of analytics is the most popular method, this method is used by 42.5% of the company. 36 % of companies use a decentralized form of analytics, with teams of analysts integrated into several or all departments of the company. And the least popular way to organize data analytics, although it is more advanced, is a mixed form of analytics organization, this form is used by 20 % of the company.

About 60 % of the company has developed regulations describing the data analysis procedure, as well as developed a clear reporting system for employees engaged in data analysis. Another important aspect of the organization of data analytics is who initiates the use of data analytics in companies. Analyzing the results, we can see that in 65% of the company, the initiator of the use of analytics is the top management, in 23 % of the companies, the initiative came from the middle management. And only in 10 % of companies, the initiative came from employees not from the company's management. Another important organizational indicator is who took responsibility for the data analysis process. So, we can see that in 37 % of companies, a new position was created to manage analytics, and in 26 % of companies, the management of analytics was taken over by the heads of departments in which analytics is implemented. The CEO of the company took responsibility for data analytics processes only in 16 % of the companies.

To test the hypotheses formulated earlier, a backward stepwise regression analysis was performed. The dependent variable was how analytics is used in a company. All other variables were selected as independent variables. In total, two previously formulated hypotheses about the company's analytics goals were tested:

H3: How analytics is organized in a company depends on how widely analytics is used

H4: How analytics is organized in a company depends on how many data analysts work in the company.

We can see from Table 5 that the way a company organizes data analytics depends on five variables. Among these variables, there are variables that were used in the hypotheses. We see that the way analytics are organized in a company depends on how widely analytics are used in the company, the coefficient is 0.329. To tell us that the more widely a company uses analytics, the more likely it is that the company uses a mixed way of organizing analytics. We also see that the way analytics are organized depends on how many data analysts work in the company, the coefficient is -0.341. This tells us that the more data analysts work in a company, the more likely it is that the company uses mixed organization analytics. The coefficient is negative because the answer options about the number of working analysts of the company were encoded in ascending order. Thus, we can confirm both hypotheses.

In addition to the two variables defined by the hypothesis, the way a company organizes analytics also depends on three other variables, such as how data analytics is funded, whether there is a reporting system for employees involved in data analysis, and whether there are plans for what data should be collected. This tells us that with the introduction of more advanced methods of organizing analytics, such as mixed methods, the company has more standards and protocols in the field of data analytics. The coefficient for variables of how data analytics is funded is 0.449. The coefficient is positive, which tells us that if the company uses more decentralized data analytics, then the funding also becomes more decentralized, that is, the money is allocated by each department independently.

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
.691[z]	.477	.417	.566

Coefficients[a]					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	1.561	.279		5.603	.000
The organization determines in advance what data should be collected	-.596	.187	-.314	-3.190	.002
How widely is data analytics used in your company?	.329	.149	.223	2.216	.030
Does the company have a clear reporting system for employees involved in data analysis?	-.444	.191	-.279	-2.320	.024
How is the company's data analysis funded?	.449	.124	.359	3.619	.001
How many data analysts are currently employed by your company?	-.341	.078	-.514	-4.391	.000

a. Dependent Variable: How is data analytics organized in your company?

Table 5. Regression results for organizational issues

4.1.3. Target perspective

Another important aspect of analytics is how the company plans its data analysis activities. 86 % of respondents participating in the survey agreed that the company's goals in analytics correspond to the company's strategic and tactical goals. Also, 70 % of the company has long-term planning and short-term planning of activities in the field of analytics.

Analyzing the company's expectations from the implementation of data analytics and the results obtained, we can see that the company's expectations do not always become real. So, we see that 53 companies planned to gain the competitive advantages of implementing analytics, and only 48 companies actually managed to do this. Also, 36 companies planned to increase customer loyalty, but only 30 managed to do it. However, some companies managed to gain an advantage, although they did not expect it. So, we see that 65 companies planned to save money and increase efficiency, and 66 companies managed to do this. Also, 54 companies planned to open new business opportunities, but 59 companies managed to do so. In general, we can see that in most cases, companies achieve fewer benefits than they initially plan.

To test the hypotheses formulated earlier, a backward stepwise regression analysis was performed. The dependent variable was the sum of the benefits received by the company after the implementation of data analytics began. All other variables were selected as independent variables. In total, two previously formulated hypotheses about the company's analytics goals were tested:

H5: How much benefit a company gets from implementing data analytics depends on how it plans data analytics

H6: How much advantage a company will gain from implementing data analytics depends on how many goals the company had before implementing data analytics

From Table 6, we can see that the number of benefits obtained from implementing data analytics depends on six variables, with a significance level of less than 0.05. We see that among the variables there are two variables that reflect how the company plans its data analytics, this is both a change about long-term analytics planning, and long-term analytics planning. The coefficients for these variables are -1.102 and -1.198, respectively. This means that those companies that have such types of planning as a result have received more benefits from the implementation of data analytics. Thus, we can conclude that the fifth hypothesis is confirmed. Already among the independent variables, there is a variable that reflects how many benefits the company planned to get from the implementation of analytics before the start of its implementation. The coefficient is 0.129, which means that the more the company planned to gain an advantage, the more it gained an advantage in the end. Thus, we can conclude that the sixth hypothesis is confirmed.

We can also see from the results that three other variables affect how many benefits a company has received from implementing data analytics. We see that the earlier the company started implementing data analytics, the more benefits it received from its implementation. This may suggest that it takes time to get real benefits. Also, the other two variables are the sum of the data types and the sum of the analytics types used in the company. This tells us that in order to get more benefits, the company needs to try to collect as many different types of data as possible, and use a wide range of types of data analytics, not only descriptive analytics, but also predictive and prescriptive.

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
.851	.723	.701	1.247

Coefficients[a]					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-1.605	.694		-2.312	.024
How long ago did the company start implementing data analytics methods?	1.071	.202	.386	5.312	.000
The company has a short-term planning of analytical work (Day-Month)	-1.102	.260	-.286	-4.230	.000
The company has a long-term planning of analytical work (several months - Several years)	-1.198	.387	.233	3.095	.003
sum_of_data_types	.288	.052	.478	5.503	.000
sum_of_types_of_analytics	.503	.193	.228	2.608	.011
sum_Why_invested	.129	.073	.135	1.763	.008

a. Dependent Variable: Sum_Advantages_gained

Table 6. Regression results for target issues

4.1.4. Analyst perspective

Another important aspect of the effectiveness of the company's data analytics application is how companies manage the competencies of a data analytics specialist. Analyzing the results, we can see that 32 % of the company employs from 31 to 100 data analysts and 30 % of the company employs more than 100 data analysts. 21 % of companies employ one to 10 data analysts and 16 % of companies employ 11 to 30 data analysts. Thus, we see that 60 % of the company employs more than 30 people. At the same time, we see that 85% of companies continue to increase the number of employees employed in the field of data analysis. This adds complexity to the company's ability to manage the competencies of these employees.

The result of the survey shows that in general, the company effectively manages the competencies of employees engaged in data analytics, so 66% of respondents agree with this. At the same time, only 42.5% of respondents agreed that their company lacks competence in any field of data analysis. In 74 % of companies, they strive to constantly improve the competence of data analysts.

To test the hypotheses formulated earlier, a backward stepwise regression analysis was performed. The dependent variable was how effectively the company manages the competencies of data analysts. All other variables were selected as independent variables. In total, two previously formulated hypotheses about the company's analytics goals were tested:

H7: How effectively a company manages data analytics competencies depends on how widely the company applies data analytics

H8: How effectively a company manages the competencies of data analysts depends on how many data analysts work in the company

From Table 7, you can see that the effectiveness of managing data analytics competencies depends on five variables. The coefficient for the variable reflecting the breadth of use of the company's data analytics is 0.181. This tells us that the more widely a company uses data analytics, the more effectively it manages the competencies of its analysts. Based on this, we can conclude that hypothesis number seven is confirmed. We also see that the effectiveness of managing the competencies of data analysts depends on how many data analysts work in the company. The coefficient for this variable is -0.141. This tells us that the more data analysts a company employs, the more effectively the company manages their competencies. Accordingly, we can conclude that the eighth hypothesis is also confirmed. All of the above tells us that with the increase in the use of data analytics, and with the increase in the size of analytical departments, companies are beginning to take a more responsible approach to managing the competencies of data analysts.

Among other variables that affect how effectively a company manages the competencies of data analysts, we see a variable that reflects the company's desire to hire the best specialists in the field of data analytics. The coefficient for this variable is 0.285, which tells us that companies that are looking for more competent data analysts are better able to manage the competencies of data analysts. We also see that those companies that increase the cost of data analytics are also more advanced in managing the competencies of data analysts. The coefficient for this variable is -0.195. We also see that companies that are better managed by the competence of data analysts do not lack competence in different areas of data analytics. The coefficient for this variable is -0.218.

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
.704	.495	.439	.352

Coefficients[a]					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-.296	.179		-1.653	.103
Have your analytics costs increased or decreased compared to previous years?	-.195	.083	-.254	-2.354	.022
How widely is data analytics used in your company?	.181	.078	.292	2.313	.024
The company is constantly looking for the best specialists in data analysis	.285	.106	.281	2.682	.009
The company lacks expertise in one or more areas of data analysis	-.218	.095	-.232	-2.287	.026
How many data analysts are currently employed by your company?	-.141	.051	.316	2.777	.007

a. Dependent Variable: The company effectively manages the competencies of data analysts

Table 7. Regression results for analyst issues

4.1.5. Technologies perspective

Analyzing what tools and methods companies use for data analysis, we see that 99 % of companies use various programming languages for data analysis such as Python or R and libraries for them. Also, 81 % of the company uses Excel spreadsheets, and 86 % of the companies use Bi systems. Less popular tools are programs like Rapid Miner and statistical programs like SPSS. Analyzing what types of analytics companies use, we see that about 85 % of companies use descriptive, diagnostic or predictive analytics. And only 63 % of companies use prescriptive analytics.

In terms of data analysis methods, we see that the initial stages of data processing, such as data cleaning, data integration from multiple sources, and error detection, are used by about 95 % of companies. As for data analysis algorithms we see that more complex methods such as deep learning algorithms neural networks are used more often than simpler methods such as factor analysis or cluster analysis. Thus, neural networks are used by about 80 % of companies, while factor analysis is used by only 68 % of companies. From this we can conclude that companies tend to use more complex methods of data analysis while neglecting simpler methods of data analysis.

To test the hypotheses formulated earlier, a backward stepwise regression analysis was performed. The dependent variable was the sum of the data analysis methods used by the company. All other variables were selected as independent variables. In total, one previously formulated hypothesis about the company's analytics goals was tested:

H9: The number of data analysis methods used depends on the width of the analytics usage

From the results of the regression analysis presented in Table 8, we can see that the number of data analysis methods used depends on four variables with a significance level less than 0.05. The number of data analysis methods used depends on how widely the company uses data analytics, the coefficient for this variable is -2.085. This tells us that the more companies use data analytics, the more they use data analysis techniques. Thus, we can conclude that the ninth hypothesis is confirmed.

Also in the results, we can see that the number of data analysis methods used also depends on how long the company has been using data analytics. The coefficient for this variable is 0.985. This tells us that the longer a company applies data analytics, the more data analysis methods it uses. Also, the number of methods used for data analysis depends on the number of types of analytics used. The more different types of analytics a company uses, the more methods are used. The number of data analysis tools used also affects how many data analysis methods are used, with a coefficient of 0.777.

Model Summary			
R	R Square	Adjusted R Square	Std. Error of the Estimate
.767[e]	.589	.560	2.718

Coefficients[a]					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	8.755	1.905		4.596	.000
How long ago did the company start implementing advanced data analytics methods?	.985	.425	.198	2.317	.023
How widely is data analytics used in your company?	-2.085	.552	-.386	-3.775	.000
sum_of_types_of_analytics	1.015	.462	.256	2.196	.031
sum_of_instruments	.777	.312	.269	2.495	.015

a. Dependent Variable: Sum_of_methods

Table 8. Regression results for techniques issues

4.2. Maturity levels of analytics based on real performance of Russian companies

A cluster analysis was conducted to determine the current level of maturity of the companies and the characteristics inherent in each level. The cluster analysis was performed in the Rapid Miner program using the PAM algorithm. For the cluster analysis, 25 variables were used, including the breadth of the use of analytics in the company, the size of the analytical department and the methods used.

The companies were grouped into 4 groups. This number of clusters was selected based on the analysis of existing models of analytical maturity. As discussed in Chapter 1.6, on average, analytical maturity models define 4 to 5 maturity levels. Since among the respondents surveyed, there were no such people who work in companies that do not use analytics at all, it was decided to separate the companies into 4 maturity levels, and assume the existence of a level 5 for companies that do not use data analytics at all.

The variability of each clustering variable between clusters was evaluated using the Chi-Squared test. For most variables, p-value was < 0.001. Thus, the variability of the width of analytics usage in the company is significant among clusters at the level of <0.001. Therefore, we can consider each cluster as a separate level of analytical maturity.

We can see in Figure 7 that as the level of maturity increases, the number of companies that apply analytics in all processes and departments of the company increases. So, at the first level of maturity, there are no companies that use data analytics in all departments, and at the fourth level of maturity, 100% of companies use analytics in all departments.

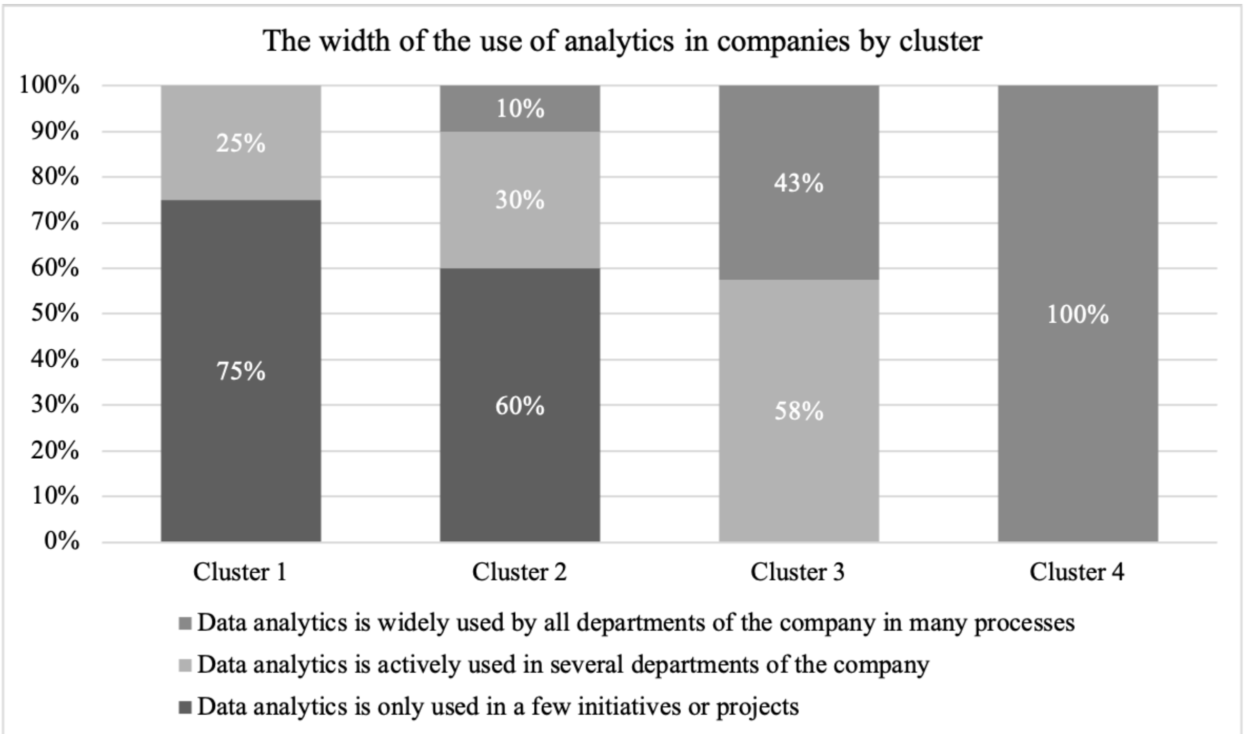

Figure 7. The width of the use of analytics in companies by cluster

Figure 8 shows the distribution of companies by cluster. The first cluster contains 15% of all companies, these companies are at the initial stage of implementing analytics. The second cluster has 12.5% of companies. The third cluster is the largest cluster with 50% of the companies. And the fourth cluster consists of 22.5% of companies, these companies are the most advanced companies in the field of analytics implementation.

Cluster

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cluster 1	12	15.0	15.0	15.0
	Cluster 2	10	12.5	12.5	27.5
	Cluster 3	40	50.0	50.0	77.5
	Cluster 4	18	22.5	22.5	100.0
	Total	80	100.0	100.0	

Figure 8. Number of companies in each cluster

In Figure 9, we can see the distribution of the data analysis methods used across the clusters. We see that with the increasing complexity of the data analysis method, the number of companies using them decreases, but we also see that the most complex data processing methods such as neural networks and deep learning methods are used more than such methods as cluster analysis or factor analysis. At the same time, there is an interesting trend that companies of the first and second cluster use such advanced data analysis methods as deep learning methods and neural networks more than simpler data processing methods such as clustering or correlation analysis. While the companies of the third and fourth cluster use such advanced methods of data analysis less or at the same level as the simpler methods of data analysis. This may indicate that the companies of the first and second cluster, despite lagging behind in other organizational areas that affect the maturity of data analysis, are trying to use advanced data analysis methods.

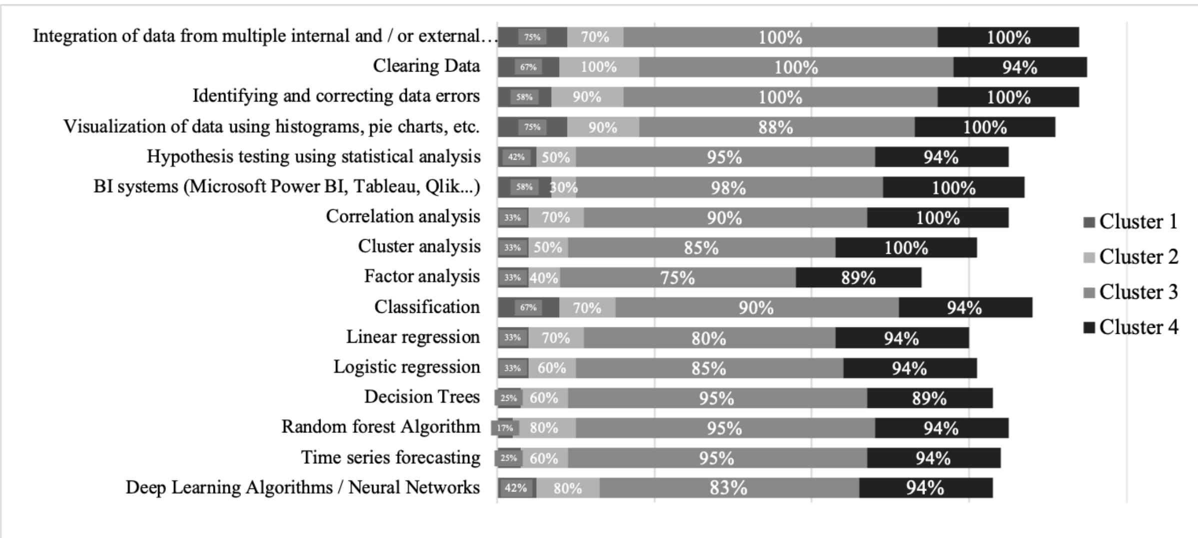

Figure 9. Using of various data analysis methods by companies from different clusters

The clusters were analyzed in descending order of analytical maturity, that is, starting from the fourth and ending with the first. The description of each cluster contains a description of the characteristics of the company in accordance with the five dimensions defined earlier: data, analytics governance, analytics goals, analysts, and data analysis methods and tools. For each cluster of companies, starting from the third, the features that distinguish companies in this cluster from more mature ones in analytics were highlighted and recommendations were given on how to move to the next cluster and become more mature in analytics.

4.2.1. Cluster 4 description

The fourth cluster consists of companies that are the most advanced in the field of data analysis. From the analysis of the results, we can see that 16.7% of companies in this cluster increased their spending on data analysis, and 83.3% of companies significantly increased their spending on data analysis over the past year. This result may indicate that these companies see a huge potential in data analytics and are willing to spend more money on it. On the Figure 7 we can see that in this cluster, all companies use data analytics in all departments and processes of the company, where possible. At the same time from Figure 10 we can see that 66.7% of the companies in this cluster use a mixed type of data analysis organization, which means that there is a central data analysis team in the company and local data analysis teams in various departments of the companies. but some companies also use other forms of data analytics organization. Thus, 22.2% of companies in this cluster use only a centralized type of data organization, with one team of data analysts for the entire company, and 11.1% of companies use only teams of data analysts that are embedded in various departments of the company. This shows that companies that are advanced in data analysis tend to organize their analytical teams in a mixed form.

88.9% of the companies in this cluster use all possible data sources, both inside and outside the company. 11.1% of respondents said that they do not know exactly about the company's data sources. This tells us that mature companies are looking for all possible data sources, none of the respondents stated that the company uses only internal or external data sources. This is confirmed by the responses of the respondents, as they all stated that the company strives to collect as much data as possible. At the same time, companies constantly update the data in their storage, as stated by 88.9% of respondents. In their companies, data is updated in real time, that is, they have implemented the collection of streaming data. Also, companies from this cluster have developed data storage rules and follow them, as stated by 88.9% of respondents. That is, companies take the data collection process seriously, which confirms the fact that the quality of the collected data is measured and regulated in all companies in this cluster. At the same time, 88.9% of respondents said that their company determines in advance what data should be collected. The result of such a responsible approach to working with data can be a well-built culture and attitude to data in the company. This is confirmed by the results of the survey, as all respondents said that their company treats data as a strategic asset. In terms of data storage, only 55.6% of companies store their data in a single data warehouse. 44.4% of companies store data in different databases. This may be the reason that only 61.1% of companies in this cluster can access data for analysis within a day or faster.

From organizational point of view, we can see that companies in this cluster have developed a clear reporting system for employees engaged in data analytics, as stated by 94.4% of respondents, and only 5.6% of respondents said that their company does not have such a reporting system. Also, in addition to a clear reporting system in analytics, 88.9% of companies have developed rules and regulations that describe the process of data collection and analysis at each stage. From the point of view of the budget allocated for analytics, everything is not so clear. 55.6% of companies in this cluster have a single dedicated budget for analytics, while in 33.3% of companies, the budget for analytics is allocated independently by each department. This is probably due to various forms of data analytics organization in companies. So, companies using a mixed form of analytics can have both a single budget for analytics allocated by the company, and project budgets for analytics allocated by different departments of the company. The initiative to implement data analytics came from the company's top management in 88.9% of companies and in 11.1% of companies, the initiative came from one of several employees from the company's management. From Figure 12 we can see that the position of Chief Analyst or Chief Data Officer was created to manage analytics in 61.1% of the companies in this cluster. In 16.7% of the companies, one of the existing chief directors took over the management of analytics. All this suggests that data analytics is top down and managed at the highest level.

All respondents stated that the company's data analysis goals correlated with the company's strategic and tactical goals. 83.3% of companies have long - term and short-term planning for data analysis. In terms of the expectations from the implementation of analytics and the results obtained, we can see on Figure 13 and Figure 14 that, in general, the results of the implementation of analytics correspond to the expectations of companies in this cluster. The most anticipated benefits for the company were cost savings and improved efficiency, increased revenue, and competitive advantages. We also see that some companies have received more results from analytics than expected, in particular, some companies have improved the company's reputation, improved brand awareness and improved customer loyalty, although they did not expect this initially.

77.8% of companies have more than 100 people employed in data analysis, and 22.2% of companies have between 31 and 100 people employed in data analysis. Thus, in all companies, the size of data analysis departments is relatively large. 16.7% of companies have increased the number of employees employed in data analysis, and 77.8% of companies in this cluster have significantly increased the number of data analysts in their data analysis department over the past year. This, along with the fact that companies are also increasing the budgets allocated to data analysis, suggests that companies see potential in data analytics. In terms of human resources management, 94.4% of respondents said that their company is constantly trying to find the best data analysts. 66.7% of respondents agreed that their company employs world-class data analysts. At the same time, only 22.2% of respondents said that the company lacks competence in several areas of data analysis. As we can see from Figure 11, the companies of the fourth cluster use all possible methods of additional education of specialists in the field of data analysis. The most common methods of employee training in this cluster are self-education and exchange of experience with the most experienced employees within the company. The majority of respondents believe that their company is committed to improving the competence of its data analysts and that their company is effectively managing its data analyst competencies, 83.33% and 88.9%, respectively

As we can see from Figure 9 almost all companies in this cluster use simple data processing methods such as data cleaning data error correction data visualization. The share of companies using more advanced data analysis methods decreases slightly with the increasing complexity of the analysis. But at the same time, approximately equal shares of companies use both the most advanced data analysis methods such as neural networks and deep learning and conventional data analysis methods such as Clustering or classification. This suggests that companies try to take full advantage of any data analysis methods and do not focus only on the most advanced data analysis methods. Additionally, 88.9% of the companies in this cluster use real-time data analytics. 83.3% of the companies in this cluster implement their analytical models in the company's client and enterprise applications. Which suggests that these companies are quite advanced in data analytics and are trying to use all the features of data analytics.

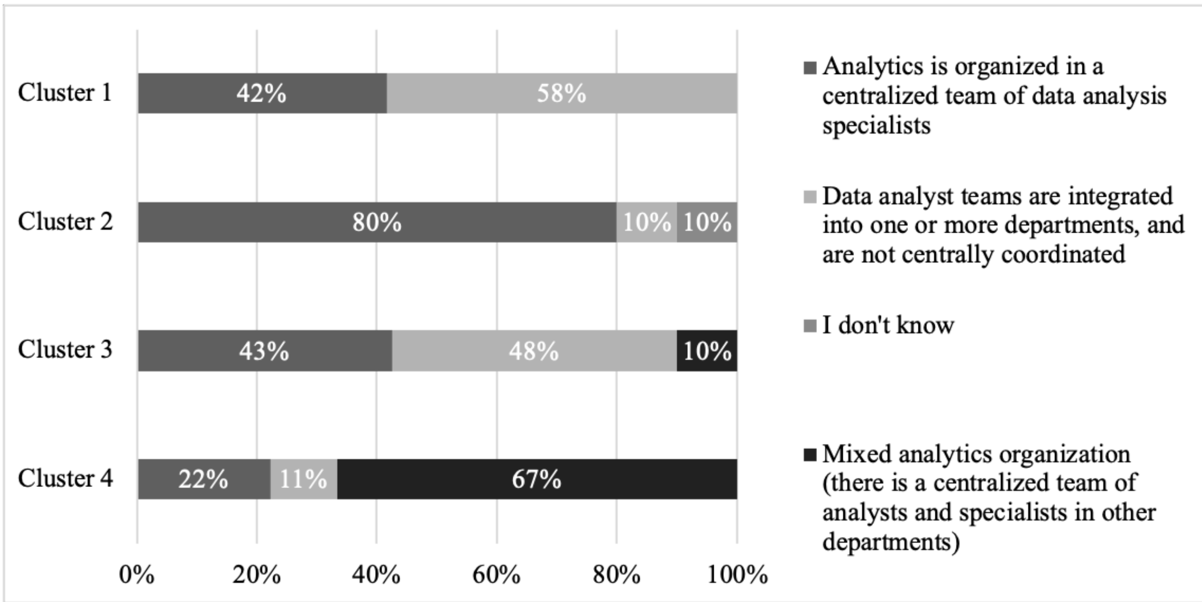

Figure 10. Ways to organize data analysis by cluster

4.2.2. Cluster 3 description

The third cluster contains companies that are slightly less advanced in implementing data analytics. This is the largest cluster containing 40 companies, which is half of all the companies surveyed. In contrast to the fourth cluster, the number of companies that significantly increased spending on data analytics is less and is 65 %. However, the number of companies that slightly increased spending on data analytics is more than in the fourth cluster and is 32.5%. This suggests that in contrast to the fourth cluster, companies in the third cluster also increase the cost of analytics, but they do not do it so actively. In contrast to the fourth cluster, we can see that this cluster contains not only companies that use analytics in all departments and processes of the company, but also 57.5 % of companies that use analytics in only a few departments of the company. Also in Figure 10, we can see that the way analytics are organized is also different from the fourth cluster, as only 10 % of companies use a mixed analytics organization system. 47.5 % of the companies in this cluster use teams of analysts integrated into the company's departments. From this, we can conclude that due to the fact that companies in this cluster, unlike companies in the fourth cluster, use analytics in fewer processes, they prefer to organize their analytics either in centralized teams or directly in departments that use analytics.

In terms of working with data, we can see that 92.5 % of the companies in this cluster use all possible data sources both inside and outside the company. However, unlike the fourth Cluster in this cluster, 5 % of the company said that they only use data collected within the company. In general, the desire to collect as much data as possible in this cluster is lower than in the fourth cluster, as 15 % of respondents said that their company does not seek to collect as much data as possible. In addition, in this cluster, only 77.5 % of companies determine in advance what data should be collected. We also see that in this cluster, the number of companies that update data continuously, that is, every second, is less than in the fourth cluster and is 75 %. While 20 % of companies said that they update the data daily or several times a week. In contrast to the fourth cluster, this cluster contains 10 % of companies that do not evaluate and do not regulate the quality of the collected data, whereas in the fourth cluster, all companies do this. From the point of view of data storage in this cluster, the company has developed less data storage rules. Also, fewer companies store data in a single data warehouse. Therefore, the share in the company in which all divisions of the company can access data within one day is 40 %, which is less than in the fourth cluster. In general, we can see that the share of respondents who believe that their company treats data as a strategic asset is 90 %, which is less than in the fourth cluster.

From the point of view of the organization of analytics, we can see that the process of analytical work in the company of the third Cluster is less organized than in the companies of the fourth cluster. So, we see that 30 % of respondents said that the company does not have a clear reporting system for employees engaged in data analysis. Also, a smaller proportion of companies have developed regulations and rules that describe the process of data analysis at all stages. Funding analytics in this cluster corresponds to the ways of organizing analytics, so 42.5 % of companies have a single dedicated budget for analytics, and in 45 % of companies, analytics is funded by each department independently. From the point of view of the analytics implementation initiative, we see that the percentage of companies in which the initiative came from the top management of the company decreased to 70 %, while in 20% of companies the initiative came from middle managers. This trend tells us that in less mature companies, support for analytics from senior management is not at the same high level as in more mature companies. From Figure 12 we can see that we also see that only 37.5 % of the company has created a new position for managing analytics at the highest level. In other companies, the functions of managing analytics have been taken over by either one of the existing directors or the manager of the department where analytics is implemented.

In terms of the company's goals in analytics, the company in this cluster is similar to the company in the fourth cluster. 95 % of respondents stated that they believe that the goals and objectives of the analytics company correspond to the strategic and tactical goals of the company. There is also a large share in companies that have long-term and short-term systems for planning analytical activities, but this share of the company is slightly less than in the fourth cluster, by about 10 %. From Figure 13 and Figure 14, we can see that the results obtained from analytics meet the expectations of companies in the third cluster. We also see that some companies did not get the desired results, as 73 % of the company said that they expected to improve their understanding of the company's processes, but only 63 % actually did it. Also, some companies did not get the expected results in improving the company's reputation and increasing customer loyalty.

Unlike the fourth cluster, this cluster has companies with small data analytics departments. So only 20 % of companies in the data analytics department have more than 100 people. In 25 % of companies, the analytics department is between 11 and 30 people in size. In 47.5 % of the companies, there are 31 to 100 people in the analytics department, which is more than in the fourth cluster. Also, in a smaller number of companies, the number of the analytics department increased significantly, but the share of companies in which the analytics department increased not significantly more than in the fourth cluster. At the same time, we see that the share of respondents who believe that their company lacks competence in one or more areas of data analysis is greater than in the fourth cluster. Also, a smaller proportion of respondents believe that the company employs world-class data analysts, only 40% of such respondents. The reason for this may be that a smaller share of the company is constantly looking for the best specialists in the field of data analysis, such companies in this cluster are 70 %. From Figure 11, we can see that in the third cluster, as in the fourth, companies use all possible methods of additional training of employees in the field of data analytics. However, we can see that in this cluster, 10 % said that their company is not engaged in additional employee education. We also see that in contrast to the fourth cluster, a smaller share of the company uses cooperation with universities in the field of additional education of employees. Also, a large proportion of companies use self-education of employees as a method of professional development. These results may indicate that companies in this cluster are less interested in providing additional education for employees. Overall, 14 % fewer respondents in this cluster believe that their company effectively manages the competencies of data analysts than in the fourth cluster.

In the companies of the third cluster, there are approximately the same trends as in the companies of the fourth cluster. Almost all companies use simple data processing methods, but with the increasing complexity of the data analysis method, the share of companies using them decreases. We also see that the share of companies using advanced data analysis methods remains approximately the same regardless of the method and is on average equal to 80 %. This tells us that companies do not seek to use more complex data analysis methods such as neural networks, and try to get all possible benefits from simpler data analysis methods. In terms of implemented technologies, in contrast to the fourth cluster, where there were no respondents who said that they did not have an analysis of the flow of collected data in the third cluster, 32.5% answered so. Also in this cluster, only 67.5 % of respondents said that analytical models are implemented in the company's corporate client applications.

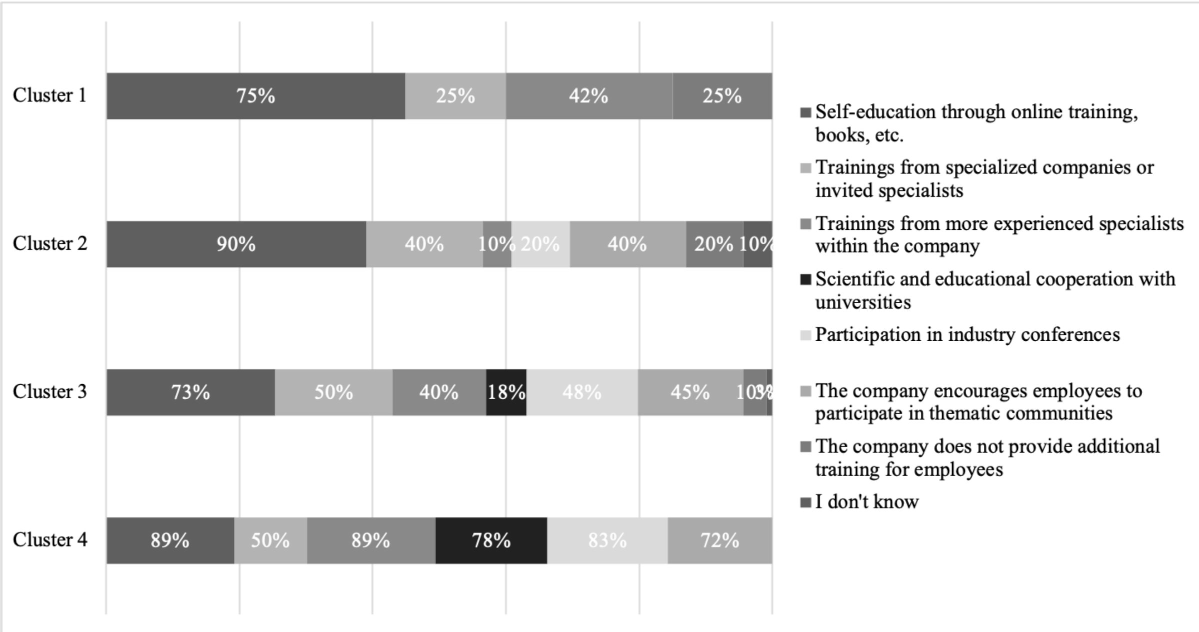

Figure 11. Methods of training analytics specialists used by companies from different clusters.

4.2.3. Cluster 2 description

The second cluster is the smallest and includes 10 companies, i.e., 12.5 %. The company in this cluster uses analytics less than companies in more mature Clusters, so only 10 % of the company uses analytics in all departments of the company, 60 % of the companies in this Cluster use analytics only within a few initiatives. 30 % of companies use analytics only in a few departments of the company. From the point of view of the organization of analytics, we see that the company does not use a mixed form of organization of analytics in most companies, but rather in 80 %, analytics is organized in centralized teams of data analysis specialists and in 10% of companies, analytics is organized in one or more departments. 60 % of companies in this cluster have significantly increased their spending on data analytics over the past year, but this is less than companies in other more mature clusters. 30 % of companies said that expenses increased but only slightly.

In terms of working with data, we see that 70 % of the companies in this cluster use all possible data sources, but compared to more mature companies, 30 % of the companies in this cluster collect data only from internal sources. This information is supported by the fact that 70 %

of respondents agreed that their company is committed to collecting as much data as possible, 30 % of respondents said that their company is not committed to collecting as much data as possible. Also, only 70% of companies in this cluster assess and regulate the quality of data in advance, 20% said that the quality is not evaluated and not regulated. Also, only 60 % of this cluster has data storage rules that companies follow, which is much less than in more mature companies. Only 60 % of respondents agreed that their companies treat data as a strategic asset, which is 30% less than in the third more mature cluster.

From the point of view of data analytics management in the company, only 60 % of the companies in this cluster have a clear reporting system for employees engaged in data analysis, which is less than in more mature companies. Also, only 50 % of companies have regulations and rules that describe the process of data collection and analysis at all stages. Funding for data analytics in 50% of companies is provided by each department independently, and only 40 % of companies have a dedicated budget for data analytics. Compared to more mature companies, the financing of analytics is more departmental. The most significant differences in this cluster are in the area of data analytics initiatives. Only in 50 % of companies, the initiative to implement data analytics came from the company's top management. Which is significantly less than in more mature companies. In 40% of the company, the initiative came from middle managers. This trend is reflected in who is responsible for analytics. From Figure 12 we can see that only 30 % of companies have created a new position for managing analytics. In 50% of the company, the responsibility for managing data analytics was taken by the heads of departments where analytics is used.

In terms of the goals of a data analytics company, this cluster is very different from clusters with more mature companies. So only 80 % of respondents agree that the goals and objectives of the company in analytics correspond to the strategic and tactical goals of the company, 20 % of the company do not agree with this. Also, in the companies of this cluster, the situation with planning is slightly worse, so only 70 % of respondents agreed that their companies have short-term planning, 60 % agreed that they have long-term planning. From Figure 13 and Figure 14, we can see that the results obtained from analytics do not meet the company's expectations. So, we see that 70 % of companies expected to gain competitive advantages from using analytics, but only 20 % of companies actually received these advantages. Also, 70 % of the company expected to get new business opportunities and only 60 % actually got them. However, only 50 % of companies expected to reduce costs and increase efficiency, but in reality, 60% of companies managed to do this.

Companies in this cluster are characterized by their small number, so 40 % of companies have an analytics department consisting of from one to 10 people. In 30 % of companies, the analytics department employs between 31 and 100 people. However, there are also large analytics departments in this cluster. In 20 % of the companies in this cluster, the analytics department employs more than 100 people. However, we see that 60 % of the companies in this cluster have significantly increased the number of employees employed in analytics over the past year. This may indicate that the company sees potential in this area and plans to develop in analytics. 60 % of respondents agree that the company is constantly looking for the best specialists in data analysis. However, only 20 % of respondents agreed that their company employs world-class data analysts. At the same time, 50% of respondents agree that their company lacks competence in one or more areas of data analysis. From Figure 11, we can see that companies in this cluster mainly use self-education as a method of improving the skills of their employees in data analysis. Also, 20 % of respondents expressed the opinion that their company does not provide opportunities for additional education. In addition, companies in this cluster do not use partnerships with universities in the field of education of their employees. Also, significantly fewer companies encourage participation in industry conferences for their employees, as opposed to more mature companies. In general, we can conclude from the results that the company is not ready to spend a lot of resources for additional education of its employees. Also, all of the above is confirmed by the fact that only 40 % of respondents agreed that their companies successfully manage the competencies of data analysts.

In Figure 9, we can see that in the company of the second cluster, in contrast to the company of the third cluster, we see a tendency to reduce the use of even the simplest data processing methods. We also see trends that more complex methods of data analysis and processing are used less than simpler ones. However, in this cluster of companies, we also see a tendency for more sophisticated data analysis methods to be used more than simpler data analysis methods. So, 80 % said that they use deep learning methods and random forest algorithms. While the share of companies using cluster and factor analysis does not exceed 50 %. This trend tells us that companies in this cluster are trying to use as complex data analysis methods as possible while not using all the possible advantages of using simpler data analysis methods. From the point of view of the technologies used in the companies of this cluster, only 40% of respondents said that their company has implemented streaming data analytics. However, 60 % of the company uses analytical models in the company's corporate and client services and applications.

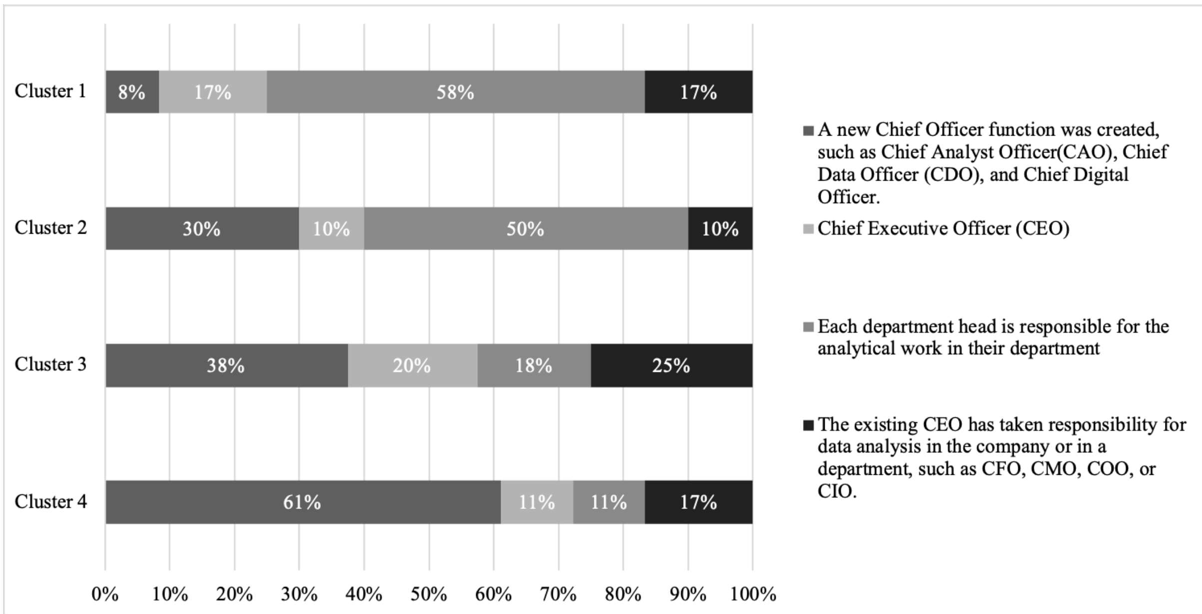

Figure 12. Who is responsible for analytics by clusters.

4.2.4. Cluster 1 description

In this cluster only 33.3% of respondents said that the company's spending on analytics has increased significantly, while 41.7% of companies do not change their spending on analytics. In other clusters, the share of companies in which the cost of analytics increased more, based on this, we can conclude that the company is not ready to increase the cost of data analytics.

In terms of data management, 50% of companies in this cluster collect data only from internal sources, and only 41.7% use all possible data sources. 50% of respondents said that their data is updated once a day or several times during the week and only 33.3% of companies in this cluster use streaming data, which means that their data is updated continuously every second. Unlike other clusters, only 41.7% of the companies in this cluster have developed rules for storing information and follow these rules. At the same time, only 25% of companies store their data in a single data warehouse. Only 41.7% of respondents in this cluster said that the company treats data as a strategic asset.

From an organizational point of view, we see that 58.3% of companies in this cluster organize their analytical teams separately in each department, and 41.7% of companies have centralized analytical teams. The reason for this is the fact that 75% of companies in this cluster use analytics only for a few initiatives or projects, so teams of analysts are integrated into the departments that lead these projects. Therefore, we also see that in 58.3% of companies in this cluster, the heads of departments are responsible for data analytics, and only in 8.3% of companies a new position was created to manage analytics. Funding analytical projects is also the task of departments, so in 66.7% of companies, data analysis is funded by each department independently, and only 25% of companies have a common budget for data analytics. All of the above is a consequence of the fact that only in 25% of companies the initiative to implement analytics came from the top management of the company, in 33.3% of companies, analytics was implemented on the initiative of employees not from the management of the companies, that is, it was the initiative of the company's employees.

Half of the respondents said that their company's goals in analytics do not correspond to the company's strategic and tactical goals. Only 41.7% of companies have long - term and short-term planning for analytics. From Figure 13 and Figure 14, we can see that in general, the results of the implementation of analytics meet the company's expectations. Some companies even managed to get more results than they expected, so only 58% of the company expected to get new business opportunities and as a result, 75% of the companies managed to do it. Also, 42 % of companies expected an improved understanding of business processes, and 58% of companies managed to achieve this. However, 33 % of the company expected an improvement in customer loyalty from the introduction of analytics, but none of the companies managed to achieve this. Thus, from the analysis, we can conclude that there is no clear relationship between expectations and the results obtained in the companies of this cluster. This may indicate that the understanding of analytics goals and how to achieve these goals is not sufficiently developed in these companies, and may lead to random results.

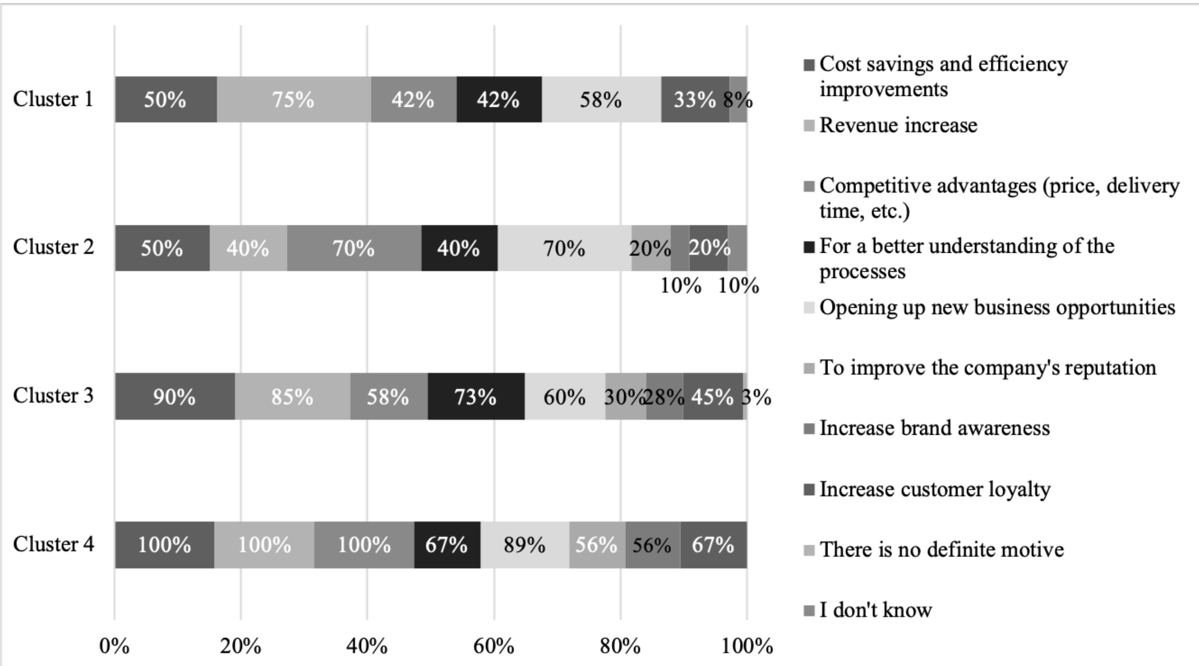

Figure 13. Why companies from different clusters invested in analytics

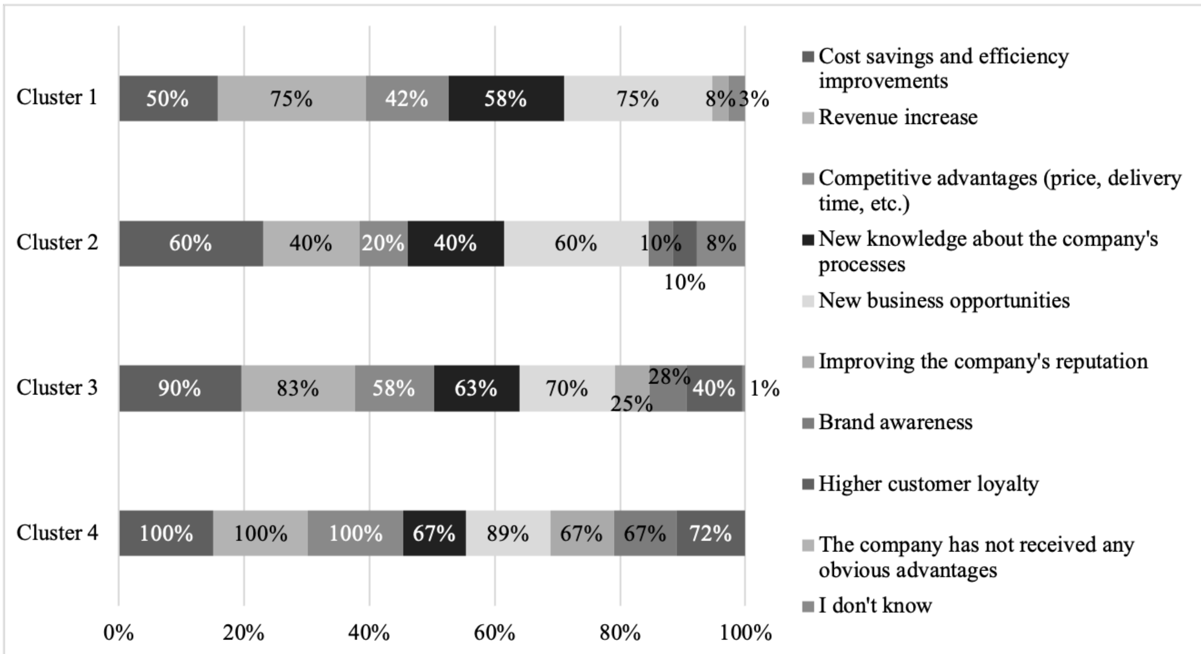

Figure 14. What benefits did companies get after implementing data analytics?

From the results of the survey, we can see that the analytical teams in the companies of this cluster are small, in 83.3% of the companies there are from 1 to 10 data analysts. At the same time, such small teams of analysts are not able to solve some tasks, 75% of respondents agreed that they lack competence in several areas of data analysis. The companies in this cluster does not try to find the best specialists in the field of data analysis, 75% of respondents agree with this. Also, none of the respondents believe that the company employs world-class data analysts. From Figure 11, we can see that companies in the first cluster do not use many ways to further educate their employees in data analysis. In 75% of companies, self-education is the main method of additional education of employees. 42 % of the company uses the transfer of experience from more experienced employees in the field of data analysis. 25 % of respondents expressed the opinion that their company is not engaged in additional education of employees. From this we can conclude that the company is not ready to spend additional resources on improving the skills of its employees. Only 25% of respondents agree that their company effectively manages the competencies of its data analysts, only 41.7% of companies seek to improve the competence of their data analysts.

From Figure 9, we can see that the companies in the first cluster have the same trends as the companies in the second cluster. So, the number of companies using more complex methods of data analysis decreases as the complexity of these methods increases. At the same time, companies tend to use more complex data analysis methods more often than simpler data analysis methods, so 42% of companies said that they use deep learning algorithms, while only 33% of companies use cluster or factor analysis. We also see that a large proportion of companies use classification analysis is 67%, which makes this method of data analysis the most popular in this cluster. Such trends tell us that companies in this Cluster tend to use the most advanced methods of data analysis, while underestimating the simpler methods of data analysis, and do not get all the benefits from using them. Only 25% of companies have implemented streaming data analysis. And only 33.3% of companies have implemented their analytical models in client and enterprise applications, such as CRM systems and so on.

4.3. Result discussion

This section will cover the results of descriptive analysis and hypothesis testing, as well as the results of cluster analysis. The result of the descriptive analysis will be a description of how companies use data analytics. The results of the hypothesis test will be discussed and conclusions drawn. Based on the results of the cluster analysis, recommendations will be given to the company in each group on how to become more mature in data analytics. At the end, the managerial implications and the limitations of this study will be discussed. Recommendations for further research in this area will be given.

4.3.1. Results of descriptive analysis and hypotheses testing

Analyzing the results of a descriptive analysis of how Russian companies implement data analytics, we can draw several conclusions. We can conclude that the Russian companies in the majority of them are quite effective and efficient in implementing data analytics in their processes. We can see that most companies take the issue of data collection seriously and have developed regulations and rules for data collection. Also, most companies tend to evaluate the quality of the data collected. In addition to the rules that describe data collection, most of the company also has rules and regulations that apply to other stages of working with data, such as preparing, cleaning, and analyzing the collected data. From the point of view of planning analytical activities, we can see that most companies take a responsible approach to this process, they have both short-term plans and long-term plans for how they will develop in data analytics, and what results they need to get. This is confirmed by the fact that the majority of respondents agreed that the company's goals in data analytics are consistent with the company's strategic goals, and that the company is moving in the right direction. Also, an indicator that Russian companies are quite advanced in the implementation of data analytics is that all companies use programming languages for data analysis such as Python or R as the main or as one of the tools. Most companies also use simpler tools such as Excel or various BI systems.

However, in addition to the positive aspects, some problems in Russian companies and issues that are worth paying attention to were also identified. As it was found out, less than half of the surveyed companies are involved in the process of collecting unstructured data such as images, video and audio recordings. This may be due to the fact that these types of data require different approaches to storage and subsequent processing and analysis. From the point of view of organizational issues of data analytics, we see that less than half of the company uses data analytics in absolutely all processes and departments of the company. This directly entails how companies organize their analytics department, few companies use a mixed type of data analytics organization. The reason for this may be that only in a little more than half of the companies, the initiator of the introduction of analytics was the top management, the CEO or some senior director. This may indicate a low level of attention to analytics issues on the part of senior management. This is confirmed by the fact that only 37 % of companies have created a new position for managing data analytics at the highest level, in other companies, one of the existing employees has taken over the functions of managing analytics. Low attention from the top management of the company to data analytics can cause slow implementation of data analytics in all processes of the company. There are also some problems in terms of working with the staff employed in the field of data analytics. Despite the fact that most companies stated that their company successfully manages the competencies of data analysts, only a few companies agreed that they have all the competencies in the field of data analysis. This suggests that there are areas of data analytics in which the company needs more specialists at a higher level. Even the fact that most companies strive to improve the competencies of data analysts does not help to improve the situation. The results also show that Russian companies tend to use as sophisticated data analysis methods as possible, such as neural networks and artificial intelligence. At the same time, without using simpler methods such as regression analysis and others. This is an area that companies that want to use all the features of data analytics need to pay close attention to.

As a result of the hypothesis test, all hypotheses were confirmed. This may indicate that Russian companies use data analytics in accordance with the existing theoretical concepts of how companies use data analytics. Since the existing literature on how companies use data analytics is mainly based on the research of foreign companies, the confirmation of the hypothesis will indicate that the use of data analytics in Russian companies in general does not differ from how foreign companies use data analytics. This may also indicate that the situation with the use of data analytics in Russian companies is most likely the same as in foreign companies. From this we can conclude that the existing models of the company's analytical maturity will be effective for Russian companies. However, despite this, the development of an analytical maturity model for Russian companies has not become less relevant. Models of analytical maturity require constant updating, taking into account the emergence of new technologies and methods of data processing. (Muller & Hart, 2016) Thus, in the future, it is necessary to consider the possibility of developing a new analytical maturity model that takes into account new trends in data analytics and is initially adapted to Russian conditions.

4.3.2. Results of cluster analysis and recommendations

As a result of the cluster analysis, recommendations were formulated for all four clusters of companies. For the fourth cluster, general recommendations were formulated, since it was not possible for companies in this cluster to formulate recommendations based on information about companies from a more advanced cluster. For all other clusters, the recommendations were given taking into account what characteristics the company has in a more advanced cluster. Thus, the recommendations formulated describe the areas in which improvement is necessary in order to be at a higher level of maturity of data analytics.

Cluster 4. The companies in this cluster are the most mature in implementing data analytics. The recommendation for these companies may be to continue improving the organizational issues of analytics. Companies can continue to improve the processes associated with data analytics, including the development of protocols and rules for working with data and their further analysis. Companies in this cluster already use analytics in all processes in the company, but companies can continue to look for new opportunities to apply data analysis in the area where it is already used. These companies can more actively adopt advanced data analysis techniques, such as deep learning and neural networks, as they successfully implement simpler data analysis techniques. In areas where companies use simpler data analytics techniques, they can try more advanced data analytics techniques, and perhaps in the process they can find new opportunities for the company. Companies should also develop data storage systems and improve data availability for all business units. This may include the creation of distributed storage systems using Hadoop technology or other solutions of this type.

Cluster 3. For companies in the third cluster, the main recommendation will be to expand the scope of data analysis, in particular, to use data analysis in more departments of the company. Companies can begin to implement simple data analytics based on the use of simple data analysis tools and methods in departments and divisions of the company in which data analytics has not been used before. Perhaps in the process of implementing analytics, this company will be able to find ways to increase efficiency and improve processes in these departments. In the process of expanding the use of data analytics in the company, you need to start moving to a mixed type of analytics organization from the centralized and distributed types of analytics organization. In addition, companies can more actively implement advanced data analysis techniques, such as neural networks and deep learning algorithms. Companies can start implementing advanced data analysis methods in processes where simpler data analysis methods were previously used. This will also help companies improve processes in other areas of data analytics, starting from the data collection stage.

Cluster 2. Companies in the second cluster should consider moving from project-based data analysis to implementing data analysis across multiple departments. This requires the company to systematize the approach to data analytics, that is, to move from project thinking to considering data analytics as a permanent tool for improving processes in the company. The company needs to restructure the work of departments to make decisions based on data analytics. An important process in this transformation is the structural transformation of analytical activities, in particular, the transition from a centralized type of analytics organization to the introduction of data analysis groups in the company's divisions or even to a mixed form of analytics organization. As a continuation of the transformation, companies in this cluster need to work out the organizational issues of using data analytics. In particular, companies need to develop systems for planning data analysis activities, as well as develop standards and protocols for working with data and analyzing data at all stages. In addition, companies in this cluster should take full advantage of simpler data analysis techniques, rather than focusing only on advanced techniques such as neural networks and deep learning algorithms.

Cluster 1. The top management of the first cluster company first needs to pay more attention to analytics and show more initiative. You need to start treating data analytics as a tool that can improve the efficiency of the company and give the company advantages. A good step on the part of top management is the initiative to introduce data analytics in several departments of the company, they use data analytics only in some projects. Companies should learn to be more attentive to data analytics, and in particular start with a more attentive attitude to data, especially in the process of data collection, and consider collecting data from all possible sources. In addition, these companies should not rush to implement advanced data analysis methods and focus on simpler data analysis methods and get all the possible benefits from using them. And in the process of working with simpler ways to analyze data, learn how to properly handle data, collect it, and store it.

4.3.3. Limitations and further research

Several limitations of the study can be identified. The first limitation is the sample size of the companies under study. The sample size was not very large and amounted to 80 companies, which is enough for an initial study of the situation with data analytics in Russian companies, but not enough in the case of developing a full-fledged model of analytical maturity adapted for Russian companies. This study also uses the opinions of employees working in Russian companies to evaluate the use of data analytics in particular company. This approach may not fully reflect the actual position of the data analytics in particular company. Thus, in future studies, it is necessary to investigate in more detail the real situation with data analytics in particular Russian companies, using other approaches. Also, in the future development of the analytical maturity model for Russian companies, it is worth paying more attention to medium and small companies, while this study focused mainly on large Russian companies.

The main limitation of this study is that the results do not offer any tools for assessing the analytical maturity of companies. This study focused mainly on the description of the current situation with analytics in Russian companies, and the search for features of the use of data analytics in Russian companies. To this end, a cluster analysis was carried out and four levels of analytical maturity were identified. However, the study does not offer any tools for companies to determine exactly what level of analytical maturity they are at. Therefore, the main focus of future research should be a deeper study of the aspects of the use of data analytics in Russian companies, and the development of a full-fledged model of analytical maturity, including tools for companies that will not be able to accurately determine at what level of analytical maturity they are. The future model of analytical maturity for Russian companies will be an important tool for the conscious development of data analytics in Russian companies.

CONCLUSION

This study was designed to explore how Russian companies use data analytics, and to explore issues related to the use of analytical maturity models in Russian companies. As a result of the study, all the research questions were answered, and important conclusions were made both from a scientific and from a managerial point of view.

At the beginning of the study, a broad review of the literature was conducted on the topic of data analytics and the topic of analytical maturity models. We found and identified problems related to the models of analytical maturity, or rather the need to update them in accordance with new technologies, as well as the need to develop models of analytical maturity adapted to the conditions of the market in which these models are used. The research questions and the purpose of the study were formulated. In order to answer the research questions, additional research hypotheses were formulated. Confirmation or refutation of the hypotheses was aimed at finding out whether Russian companies have any differences in the use of data analytics from foreign companies.

To answer these questions, a structured survey was conducted. Further, the survey results were analyzed by three different types of analysis. At the beginning, a descriptive analysis was performed, describing how Russian companies use data analytics. In the future, a regression analysis was performed in order to confirm or refute the proposed research hypotheses. And at the third stage, a cluster analysis was performed in order to divide the companies participating in the study into four groups according to the level of analytical maturity. A description of the specific features of the companies in each group was also made. As a result, all the research questions were answered.

RQ1 How do Russian companies currently use data analytics?

Answering the first question of the study, a descriptive analysis of the survey results was conducted. It was described how Russian companies apply data analytics for each of the criteria important for the analytical maturity of the company. As a result, it became clear that Russian companies take a responsible approach to working with data, most companies strive to collect as much data as possible. Also, most companies have developed rules and standards for working with data, in particular data collection and storage. Most companies take a responsible approach to planning their data analytics activities, and have short-and long-term data analytics goals. Also, all companies use advanced data analytics tools such as programming languages, in addition, most companies also use simpler tools such as Excel spreadsheets and BI systems. However, despite this, Russian companies have several problems in the application of data analytics. So few companies collect non-structured data such as images, video, or audio recordings. Also, less than half of the company uses data analytics in all processes and departments of the company. Another problem is that Russian companies tend to use complex data analysis methods such as neural networks, underestimating simpler approaches to data analysis. The main problem that potentially affects the overall situation with data analytics in Russian companies is the lack of attention to data analytics on the part of company management. Thus, as a result of the descriptive analysis, it became clear that in general, Russian companies are quite well advanced in the field of data analytics, but there are still areas in which further development is necessary.

RQ2 Can the existing models of analytical maturity be considered applicable for Russian companies?

To answer the second question of the study, research hypotheses were formulated, based on a review of the literature. For the hypothesis test, a regression analysis was performed in order to study the interaction between variables and find out the aspects that affect certain criteria for implementing data analytics of companies. As a result of the analysis, all the hypotheses put forward were confirmed, which may indicate that Russian companies use data analytics in accordance with existing theoretical ideas about how companies use data analytics. Thus, based on the conclusions made, it can be assumed that the existing models of analytical maturity are effective for use in Russian companies. However, despite this, the development of an analytical maturity model for Russian companies has not become less relevant. Models of analytical maturity require constant updating, taking into account the emergence of new technologies and methods of data processing. (Muller & Hart, 2016) Thus, in the future, it is necessary to consider the possibility of developing a new analytical maturity model that takes into account new trends in data analytics and is initially adapted to Russian conditions.

RQ3 What levels of analytics maturity can be determined for Russian companies?

To answer the third question of the study, a cluster analysis was conducted, which allowed us to divide the companies at the level according to the level of implementation of data analytics in the company data. Before that, the literature was analyzed to determine the required number of levels. Since most existing models of analytical maturity define five levels, one of which is intended for companies that do not implement data analytics, it was decided to divide the companies participating in the survey into four levels. For clustering purposes, the PAM algorithm was chosen because it gave the best results.

RQ4 What are the characteristics of companies at each maturity level?

To answer the fourth question of the study, a descriptive analysis of companies at each of the identified levels of analytical maturity was conducted. The company was described according to five criteria that affect the analytical maturity of the companies. Companies that are at the fourth level of analytical maturity are the most advanced in the implementation of data analytics among Russian companies. Data Companies responsibly approaches almost every aspect of implementing data analytics. The only aspect of implementing data analytics that these companies need to pay attention to is the organization of data storage. Few companies in this group use a single data warehouse, and a system of quick access to data for their analysis. A third-level company is less advanced in implementing data analytics than a fourth-level company. Their main difference is the lower prevalence of data analytics in the company. Also, fewer companies at this level use a mixed data analytics organization system. From the point of view of the methods used, it can be observed that companies at this level also use less advanced data analysis methods. A distinctive feature of a second-tier company is less organized data analytics. Only half of the companies in this group have standards and rules for working with data analytics. There is also a deterioration in all other criteria for implementing data analytics, such as working with data, planning analytical activities, and other organizational issues. The first-tier companies are the most lagging behind in the implementation of data analytics. These companies use data analytics mainly only for individual initiatives or projects. The company's data also shows the worst performance in other aspects of data analytics implementation. The results of this stage of the study are the basis for answering the next question of the study, that is, they allow us to formulate recommendations for the company at each level of analytical maturity of Russian companies.

RQ5 What Russian companies need to do to get to the next level of maturity?

The answer to the fifth question of the study was based on the results of the answer to the fourth question of the study. Considering the results of the descriptive analysis of the company at each of the identified levels of analytical maturity, recommendations were made for the company at each of the levels. As a result, recommendations were formulated that allow companies to understand what needs to be done in order to be at a higher level of analytical maturity. Companies at the fourth level need to continue to improve the organizational issues of analytics. More actively implement advanced data analysis techniques, develop data storage systems, and improve data availability for all business units. Companies at the third level need to expand the scope of data analysis and move to a mixed type of analytics organization from the centralized and distributed types of analytics organization. As well as more actively implement advanced data analysis methods. Companies at the second level need to make the transition from project-based data analysis to the implementation of data analysis in multiple departments. Develop a data analysis planning system, and develop standards and protocols for working with data and analyzing data at all stages. Recommendations for the company at the first level are addressed primarily to the top management of companies. Senior management should pay more attention to analytics and take more initiative. The company as a whole need to take a more responsible approach to the data collection process and collect data from all possible sources. As well as continue to improve organizational processes related to data analytics.

This research is significant both from a managerial and scientific point of view. From a scientific point of view, an important result is the proof that the use of data analytics in Russian companies is not different from the use of data analytics in foreign companies. This research is also the basis for further study of the analytical maturity of Russian companies and the development of an analytical maturity model adapted specifically for use in Russian companies. From a managerial point of view, the study comprehensively examines the aspects of using data analytics in Russian companies. This allows managers of companies that use data analytics to evaluate how their company is developing in this area. The study also contains recommendations for companies with different levels of development of data analytics, on how to improve their activities in this area.

LIST OF REFERENCES

1. Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171-209. doi:10.1007/s11036-013-0489-0
2. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A.H. Byers, Big Data: The Next Frontier for Innovation, Competition, and Productivity, McKinsey Global Institute, 2011.
3. Holst, A. (2021, February 05). Total data volume WORLDWIDE 2010-2024. Retrieved April 15, 2021, from https://www.statista.com/statistics/871513/worldwide-data-created/
4. Moed, H., 2012. The Evolution of Big Data as a Research and Scientific Topic: Overview of the Literature, 2012, ResearchTrends.
5. Laney, D. (2001). 3D Data Management: Controlling Data Volume, Velocity, and Variety. *META Group Inc.*
6. Gantz, J. and E. Reinsel. (2011). Extracting Value from Chaos, IDC's Digital Universe Study, sponsored by EMC.
7. Patgiri, R., & Ahmed, A. (2016). Big data: The v's of the game changer paradigm. *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. doi:10.1109/hpcc-smartcity-dss.2016.0014
8. Ashabi, A., Sahibuddin, S. B., & Haghighi, M. S. (2020). Big data: Current challenges and future scope. *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. doi:10.1109/iscaie47305.2020.9108826
9. Benjelloun, F., Lahcen, A. A., & Belfkih, S. (2015). An overview of big data opportunities, applications and tools. *2015 Intelligent Systems and Computer Vision (ISCV)*. doi:10.1109/isacv.2015.7105553
10. Gudivada, V. N. (2017). Data analytics. *Data Analytics for Intelligent Transportation Systems,* 31-67. doi:10.1016/b978-0-12-809715-1.00002-x
11. Maoz, M. 2013. How IT Should Deepen Big Data Analysis to Support Customer-Centricity. Gatner G00248980. Available at: https://www.gartner.com/en/documents/2531116/how-it-should-deepen-big-data-analysis-to-support-custom. Assessed 03 April 20, 2021.
12. Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Review. (Davenport & Harris, 2007)

13. Gartner (2019). Planning guide for data and analytics. Available at: https://www.gartner.com/en/doc/361501-2019-planning-guide-for-data-and-analytics. Assessed 03 April 20, 2021.
14. Duan, L., & Xiong, Y. (2015). Big data analytics and business analytics. Journal of Management Analytics, 2(1), 1-21. doi:10.1080/23270012.2015.1020891
15. Cattell, R. (2011). Scalable SQL and NoSQL data stores. ACM Sigmod Record (Special Interest Group on Management of Data), 39(4), 12–27.
16. McCreary, D., & Kelly, A. (2013). Making sense of NoSQL. Greenwich, CT: Manning Publications.
17. Carme, J., & Jiménez, F. J. R. (2011). Open source solutions for big data management. Bezons, France: AtoS.
18. Russom, P. (2013). Managing big data. Renton, WA: The Data Warehousing Institute (TDWI) Best Practices Report.
19. Dean J, Ghemawat S (2008) Mapreduce: simplified data processing on large clusters. Commun ACM 51(1):107–113
20. Apache Hadoop. (2021). Retrieved April 20, 2021, from https://hadoop.apache.org/
21. Piatetsky, G. (2019). Python leads the 11 top data science, machine learning Platforms: Trends and analysis. Retrieved April 20, 2021, from https://www.kdnuggets.com/2019/05/poll-top-data-science-machine-learning-platforms.html
22. Bonthu, S., & Hima Bindu, K. (2018). Review of leading data analytics tools. International Journal of Engineering & Technology, 7(3.31), 10. doi:10.14419/ijet.v7i3.31.18190
23. Palocsay, S. W., Markham, I. S., & Markham, S. E. (2010). Utilizing and teaching data tools in Excel for exploratory analysis. Journal of Business Research, 63(2), 191-206. doi:10.1016/j.jbusres.2009.03.008
24. Alade, A. T. (2017). Business Intelligence Tools for Informed Decision-Making: An Overview. Strategic Engineering for Cloud Computing and Big Data Analytics, 207–223. doi:10.1007/978-3-319-52491-7_12
25. Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Sethupathy, G. (2018, April 23). The age of analytics: Competing in a data-driven world. Retrieved April 20, 2021, from https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world
26. Kaur, H., & Phutela, A. (2018). Commentary upon descriptive data analytics. 2018 2nd International Conference on Inventive Systems and Control (ICISC). doi:10.1109/icisc.2018.8398884

27. Chandola, V., & Kumar, V. (2005). Summarization - compressing data into an Informative Representation. Fifth IEEE International Conference on Data Mining (ICDM'05). doi:10.1109/icdm.2005.137
28. Lepenioti, K., Bousdekis, A., Apostolou, D., & Mentzas, G. (2020). Prescriptive analytics: Literature review and research challenges. International Journal of Information Management, 50, 57-70. doi:10.1016/j.ijinfomgt.2019.04.003
29. Posavec, A. B., & Krajnovic, S. (2016). Challenges in adopting big data strategies and plans in organizations. 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). doi:10.1109/mipro.2016.7522327
30. Morabito, V. (2015). Big Data and Analytics.doi:10.1007/978-3-319-10665-6
31. Cervone, H. F. (2016). Organizational considerations initiating a big data and analytics implementation. Digital Library Perspectives, 32(3), 137–141.doi:10.1108/dlp-05-2016-0013
32. Семеновская, Е. (2019). Аналитика больших данных как инструмент бизнес инноваций. IDC при поддержке Hitachi Vantara.
33. Ассоциация Больших Данных. (2018). Стратегия развития рынка больших данных до 2024 года.
34. Król, K., & Zdonek, D. (2020). Analytics Maturity Models: An Overview. Information, 11(3), 142. doi:10.3390/info11030142
35. Ariyarathna, K., & Peter, S. (2019). Business Analytics Maturity Models: A Systematic Review of Literature. Proceedings of the International Conference on Industrial Engineering and Operations Management Bangkok.
36. Muller, L., & Hart, M. (2016). Updating business intelligence and analytics maturity models for new developments. Lecture Notes in Business Information Processing, 137-151. doi:10.1007/978-3-319-32877-5_11
37. Nda, R. M., Abdul Hamid, A., & Tasmin, R. (2020). Assessment of Big Data Analytics Maturity Models: An Overview. Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management Detroit, Michigan, USA.
38. Davenport, T. H., & Harris, J. G. (2017). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business Review Press.
39. Halper, F., & Stodder, D. (2014). TDWI analytics maturity model guide. TDWI research, 1-20.

40. Lismont, J., Vanthienen, J., Baesens, B., & Lemahieu, W. (2017). Defining analytics maturity indicators: A survey approach. International Journal of Information Management, 37(3), 114-124. doi:10.1016/j.ijinfomgt.2016.12.003
41. Kiron, D., Shockley, R., Kruschwitz, N., Finch, G., & Haydock, M. (2012). Analytics: The widening divide. MIT Sloan Management Review, 53, 1–22.
42. Davenport, T. H., Harris, J. G., & Morison, R. (2010). Analytics at work: Smarter decisions, better results. Boston, MA: Harvard Business Review Press.
43. Saxena, R., & Srinivasan, A. (2013). Business analytics: A practitioner's guide. New York, NY: Springer-Verlag.
44. LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52, 21–32.
45. СПАРК – Проверка, анализ и мониторинг компаний. Компании России - статистика организаций, каталог и рейтинг топ компаний | Полная информация о предприятиях России по отраслям на СПАРК-Интерфакс. (2021). https://www.spark-interfax.ru/statistics.

APPENDIX A. CLUSTERING VARIABLES

Variable	Chi Square significance level (p - value)
How is data analytics organized in your company?	.000
How many data analysts are currently employed by your company?	.000
How widely is data analytics used in your company?	.000
(types_of_analytics) Descriptive analytics (description of events that occurred using statistics, plotting graphs and diagrams)	.003
(types_of_analytics) Diagnostic analytics (establishing the causal relationship of events, correlation analysis, pattern detection)	.000
(types_of_analytics) Predictive analytics (determining the likely future outcome of events or the probability of situations occurring)	.000
(types_of_analytics) Prescriptive analytics (analyzing all available information in order to understand what actions should be taken to achieve a specific result)	.000
(methods) Integration of data from multiple internal and / or external sources	.001
(methods) Clearing Data	.000
(methods) Identifying and correcting data errors	.000
(methods) Visualization of data using histograms, pie charts, etc.	.035
(methods) Testing hypotheses with statistical analysis	.000
(methods) BI systems (Microsoft Power BI, Tableau, Qlik ...)	.000
(methods) Correlation analysis	.000
(methods) Cluster Analysis	.000
(methods) Factor analysis	.000
(methods) Classification	.023
(methods) Linear regression	.001
(methods) Logistic regression	.000
(methods) Decision Trees	.000
(methods) Random Forest (Random forest Algorithm)	.000
(methods) Time series forecasting	.000
(methods) Deep Learning algorithms / Neural Networks	.005

APPENDIX B. SURVEY QUESTIONS

Part 1 – Questions about company

1) What is the name of the company you work for? This data will only be used to link the responses of multiple participants in your company.
2) How long ago did the company start implementing advanced data analytics methods?

- o Less than a year ago
- o 1 - 2 years ago
- o 3 - 4 years ago
- o More than 5 years ago
- o I don't know
- o Other:

3) Does the government have a stake in your company?

- o Yes
- o No
- o I don't know

4) Does your company sell products and / or services offline, online, or both?

- o Online only
- o Offline only
- o Both online and offline
- o I don't know

5) What sector does your company operate in?

- ☐ Agriculture
- ☐ Air Company
- ☐ Audit
- ☐ Automotive business
- ☐ Biotechnologies
- ☐ Construction
- ☐ Consulting services
- ☐ Education

- Entertainment
- FMCG
- Financial services
- Food
- Gaming Industry
- Healthcare
- Hotel business
- HR services
- Information technology
- Information services
- Legal services
- Logistics
- Production
- Marketing and communication
- Mining operations
- Non-profit organization
- Online retail
- Pharmaceutical business
- Public transport
- Realty
- Researches
- Restaurants and cafes
- Retail trade
- Technologies
- Telecommunications
- Travel Agency
- Public utilities
- I don't know
- Other:

6) What is the approximate size of your company in terms of the number of employees?
7) What is the approximate annual turnover of your company? (In rubles)
8) How much did your company spend on data analytics over the past year? (In rubles)
9) Have your analytics costs increased or decreased compared to previous years?

   - Increased significantly

- Increased slightly
- Not changed
- Decreased slightly
- Decreased significantly
- I don't know

Part 2 – Questions about data

10) How does the company get the data for further analysis?

- Data is purchased from data providers
- Data is collected by the company independently from internal sources
- The company uses all available data sources both internally and externally.
- I don't know
- Other:

11) What types of data does the company collect? Check all the relevant options.

- Clickstream data (where the client clicks on a web page)
- Demographic data (age, family, income)
- Environmental data (weather, environmental pollution)
- Geographical data or geolocation data
- Market data (data on competitors, prices and offers of competitors)
- Customer data from social networks
- Data about the company's internal processes (delivery times, sales conversion, etc.)
- Feedback from customers, surveys, etc.
- Purchase history / transaction data
- Data on the customer's interaction with the product
- Data from Internet of Things (IoT) devices
- Data from sensors and sensors installed on any equipment
- Text data (emails, complaints, messages, and reviews in social networks). ad networks)
- Images (photos of products, photos of customers, etc.)
- Audio recordings (conversations with customers, etc.)
- Video (recordings from surveillance cameras, etc.)
- I do not know
- Other:

12) How often is the data in the company's databases updated?

- Data is updated continuously (every second)
- Data is updated daily (or several times a week)
- The data is updated several times a month
- The data is rarely updated (several times a year)
- I don't know
- Other:

13) The company evaluates and regulates the quality of data.

- Agree
- Disagree
- Don't know

14) The organization strives to collect as much data as possible

- Agree
- Disagree
- Don't know

15) The organization determines in advance what data should be collected.

- Agree
- Disagree
- Don't know

16) The organization has developed rules for data storage and follows them.

- Agree
- Disagree
- Don't know

17) The data for analysis is stored in a single data warehouse

- Agree
- Disagree
- Don't know

18) The company treats data as a strategic asset

- Agree
- Disagree
- Don't know

19) All divisions of the company (or employees) can quickly (within one day or faster) get access to the data

- Agree
- Disagree
- Don't know

Part 3 – Questions about managing data analysis in the company

20) How widely is data analytics used in your company?

- Data analytics is only used in a few initiatives or projects
- Data analytics is actively used in several departments of the company
- Data analytics is widely used by all departments of the company in many processes
- I don't know
- Other:

21) How is data analytics organized in your company?

- Analytics is organized in a centralized team of data analysis specialists.
- Data analyst teams are integrated into one or more departments, and are not centrally coordinated.
- Mixed analytics organization (there is a centralized team of analysts and specialists in other departments)
- Analytical work is outsourced to a third party.
- I don't know
- Other:

22) Does the company have a clear reporting system for employees involved in data analysis?

- Yes
- No

- I don't know

23) The company has regulations and rules that describe the process of data collection and analysis at all stages?

- Yes
- No
- I don't know

24) How is the company's data analysis funded?

- The company has a single dedicated budget for data analytics.
- The company does not have a single dedicated budget for data analytics, data analysis is centrally paid for as additional expenses
- Each department finances the data analysis independently, within the allocated budget
- Each department finances the data analysis independently, as an additional cost.
- There is no clear funding structure
- I don't know
- Other:

25) Who is the initiator of the implementation of data analytics in the company?

- Top management of the company (CEO, CAO, CFO, etc.)
- One or more employees from the company's management
- One or more employees not from the company's management team
- I don't know
- Other:

26) Who is the main person responsible for the data analysis processes in your company?

- Chief Executive Officer (CEO)
- An existing Chief executive officer has assumed responsibility for data analysis in the company or in a department, such as a CFO, CMO, COO, or CIO.
- A new chief Officer function was created, such as Chief Analyst (CAO), Chief Data Officer (CDO), and Chief Digital Officer.
- Each department head is responsible for the analytical work in their department
- Middle Manager (Project manager, Product manager)
- The employee is not from the company's management

- Nobody
- I don't know
- Other:

Part 4 – Questions about the company's goals in data analysis

27) In your opinion, do the company's goals and objectives in analytics correspond to the company's strategic and tactical goals?

- Yes
- No
- I don't know

28) The company has a short-term planning of analytical work (Day-Month)

- Agree
- Disagree
- Don't know

29) The company has a long-term planning of analytical work (several months - several years)

- Agree
- Disagree
- Don't know

30) Why did your company invest in analytics? Please indicate what motivations you think were most important for starting the analysis.

- Cost savings and efficiency improvements
- Increase in revenue
- Competitive advantages (price, delivery time, etc.)
- Opening up new business opportunities
- For a better understanding of the processes
- To improve the company's reputation
- Increase brand awareness
- Increase customer loyalty
- There is no definite motive
- I don't know

- Other:

31) What advantages do you think the company received after the introduction of data analysis?

- Cost savings and efficiency improvements
- Increase in revenue
- Competitive advantages (price, delivery time, etc.)
- New business opportunities
- New knowledge about the company's processes
- Improving the company's reputation
- Brand awareness
- Higher customer loyalty
- The company has not received any obvious advantages
- I don't know
- Other:

Part 5 – Questions about employees involved in data analysis

32) How many data analysts are currently employed by your company?

- 0 people
- 1-10 people
- 11-30 people
- 31-100 people
- More than 100 people
- I don't know
- Other:

33) Has the number of employees employed in data analytics increased or decreased over the past year?

- Increased significantly
- Increased slightly
- Not changed
- Decreased slightly

- Decreased significantly
- I don't know

34) How does the company train data processing specialists in new analytical methods? (Multiple answers are possible.)

- ☐ Self-education through online training, books, etc.
- ☐ Trainings from specialized companies or invited specialists
- ☐ Trainings from more experienced specialists within the company
- ☐ Scientific and educational cooperation with universities
- ☐ Participation in industry conferences
- ☐ The company encourages employees to participate in thematic communities
- ☐ The company does not provide additional training for employees
- ☐ I don't know
- ☐ Other:

35) Does the company take into account the abilities, capabilities, and desires of data analysts when allocating tasks?

- Yes
- No
- I don't know

36) The company is constantly looking for the best specialists in data analysis

- Agree
- Disagree
- Don't know

37) The company employs world-class data analysts

- Agree
- Disagree
- Don't know

38) The company strives to improve the level of competence of data analysts

- Agree

- Disagree
- Don't know

39) The company lacks expertise in one or more areas of data analysis

- Agree
- Disagree
- Don't know

40) The company effectively manages the competencies of data analysts

- Agree
- Disagree
- Don't know

Part 6 – Questions about the technologies and analytical methods used in the company

41) Which departments use data analytics?

- In the Finance Department (Analysis of financial indicators, Revenue management and pricing, etc.)
- In the Marketing department (Forecasting customer churn, Customer Segmentation, etc.)
- In the Human Resources Department (Optimization of recruitment, Employee Performance Management, etc.)
- In the Operations Department (Inventory Management, Supply chain improvement, etc.)
- I do not know
- Other:

42) What types of data analytics are used in the company?

- Descriptive analytics (describing events that occurred using statistics, plotting graphs and diagrams)
- Diagnostic analytics (establishing the causal relationship of events, correlation analysis, pattern detection)
- Predictive analytics (determining the likely future outcome of events or the probability of situations occurring)

- Prescriptive analytics (analyzing all available information in order to understand what actions should be taken to achieve a specific result)
- I don't know
- Other:

43) What tools does the company use for data analysis?

- Spreadsheets, such as Excel
- Data visualization tools, such as dashboard, BI systems
- Web analytics tools, such as Google Analytics
- Commercial analytical programs, such as SAS, SPSS
- Programs with ready-made analytical models, such as RapidMiner, Alteryx, etc.
- Python, R, or other programming languages, and libraries for them
- I don't know
- Other:

44) What data processing methods are used in the company?

- Integration of data from multiple internal and / or external sources
- Clearing data
- Identify and correct data errors
- Visualize data using histograms, pie charts, etc.
- Testing hypotheses with statistical analysis
- BI systems (Microsoft Power BI, Tableau, Qlik...)
- Correlation analysis
- Cluster analysis
- Factor analysis
- Classification
- Linear regression
- Logistic regression
- Decision Trees)
- Random forest Algorithm)
- Time series forecasting
- Deep Learning algorithms / Neural Networks
- I don't know
- Other:

45) The company has implemented the analysis of streaming data? (Analysis of the flow of collected data in real time)

- Yes
- No
- I don't know
- Other:

46) Are the created analytical models implemented in the company's client or administrative IT services (website, client applications, CRM or ERP systems, etc.)?

- Yes
- No
- I don't know
- Other:

Part 7 – Questions about the survey participant

47) What is your function in the company?

- General Director
- Senior Manager
- Business Analyst
- Data Analyst (Data Scientist)
- I don't know
- Head of the Analytics Department
- Other:

48) Do you personally work with data analytics in the company?

- No
- Yes, I have a function in analytics
- Yes, I have to base my decisions on analytics
- Yes, I work together or communicate with analysts
- I don't know
- Other:

49) Do you have any additional comments regarding the use of analytics in your particular organization or in other organizations?

50) Do you have any additional comments about this survey?

APPENDIX C. THE LIST OF COMPANIES PARTICIPATED IN SURVEY

1. Playrix
2. Циан
3. Роснефть
4. СДЭК
5. SOKOLOV
6. Siberian Wellness
7. Вайлдберриз
8. LinguaLeo
9. РЖД
10. СКБ Контур
11. Билайн
12. Grid Dynamics
13. Funbox
14. X5 Retail Group
15. Аэрофлот
16. Сбер
17. Agima
18. Ozon
19. ГК ФОРС
20. Азия Авто
21. Яндекс
22. Первый гипермаркет мебели
23. М.Видео
24. АльфаБанк
25. S7
26. НЛМК
27. ЦФТ
28. Норникель
29. Петрович
30. Ситилинк
31. Fix Price
32. Ситимобил
33. Детский Мир
34. Байкал Сервис

35. Банк Открытие
36. Самокат
37. Funbox
38. ВТБ
39. Robo Finance
40. Светофор
41. ЛЕНТА
42. Mail.ru Group
43. СовкомБанк
44. Спортмастер
45. Красное-Белое
46. АгроТерра
47. Ростелеком
48. Северсталь
49. Газпром
50. Гринатом
51. ТМК
52. Avito
53. Skyeng
54. О'Кей
55. ПИК
56. Tinkoff
57. Четыре Лапы
58. SOLOPHARM
59. Буквоед
60. Почта России
61. DNS
62. Нетология
63. Рольф
64. РОСБАНК
65. Деловые Линии
66. Азбука Вкуса
67. Протек
68. Комус
69. Дикси

70. АО "Тандер"

71. МТС

72. Лукойл

73. Технониколь

74. Unknown

75. Unknown

76. Unknown

77. Unknown

78. Unknown

79. Unknown

80. Unknown