ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

The impact of Data Analytics on audit efficiency

Master Thesis in Accounting, auditing and control

Name student: Neslihan Kalyoncu

Student ID number: 360820

Supervisor: P.Y.E Leung

Second assessor: Y. Li

Abstract

This thesis examines the impact of data analytics on audit efficiency within one of the Big Four companies in the Netherlands from 2018 until 2021. In recent years, data analytics has become a hot topic. The focus area in this thesis is the impact on audit efficiency. I find that data analytics positively impacts audit efficiency. Additionally, multiple tools or multiple times of using tools by one client also positively impact audit efficiency. The variables size, industry and the audit impact, all have a significant impact on the audit efficiency with the use of data analytics tools. Mainly Large and Medium sized firms have a positive impact on the audit efficiency. The audit impact that mainly impacts audit efficiency is the category Not determined. Because for these observations it is not determined whether it affects the audit approach, but it effects the audit efficiency positively in hours. Overall, the use of data analytics tools increases the audit efficiency. In addition, interested parties are also questioning whether the implementation of data analytics will have an impact on the employment within the audit. My findings suggest that data analytics increases audit efficiency in the audit function measured for a period of three years.

Key words: audit efficiency, data analytics, employment.

Contents

1. Introduction

Nowadays, a world without technology and without the use of Internet cannot be imagined. In the business environment a lot of use of technology and automation of tasks could be seen. The use of technology, especially data analytics is a hot topic in the audit in recent years. Data analytics has become a more significant part of the business of big accounting firms (Katz, 2014). The auditor should therefore acquire the necessary skills and knowledge to apply data analytics within the audit. A member of the Public Company Accounting Oversight Board argued that accounting firms must give a boost to the services sides of their businesses to attract analysts with data skills (Katz, 2014).

Whether the accountant will disappear in the future is the hot topic in the last couple of years within the audit. Numerous academic studies have examined data analytics in auditing, but relatively little empirical research has been conducted (Earley, 2015). According to an Oxford study, the accountant is one of the jobs that is likely to disappear in the near future as the probability of computerization is high (Frey & Osborne, 2016). Innovations such as blockchain and artificial intelligence will take over the audit work of the auditor (Bruins, 2018). According to the accountants, certainly there is no worry about the disappearance of the job as accountant. As new data systems like the blockchain will not make the accountant redundant (El Messaoudi & Kintou, 2019). The accountant still needs to control the input and output of the data and monitor the throughput (El Messaoudi & Kintou, 2019). The opinions and expectations regarding the future of the audit are different looking from an audit perspective and from an Information Technology (IT) perspective. This thesis aims to answer the research question, *"What is the impact of data analytics on audit efficiency?"* by answering two sub-questions. The first sub-question that this thesis aims to answer is *"Does the use of data analytics tools positively impact audit efficiency?"*. The second sub-question that this thesis aims to answer is *"Is the impact of data analytics tools on audit efficiency affected by size and industry?"*.

This study investigates the application of data analytics in the audit by one of the Big Four companies located in the Netherlands. Understanding how data analytics tools contribute to the audit efficiency and which factors positively impact this effect is interesting considering this debate. I expect that data analytics has a positive impact on audit efficiency, because the importance and the existence of IT in our everyday life and so also in the workplace increased in recent years. That is why we also must anticipate and react on the changes due to IT on time. I therefore expect that it facilitates our daily work if the right investments in data analytics are made.

This study uses the quantitative method and examines audit efficiency for 365 different companies over a period of three years to provide quantitative evidence for the debate on audit efficiency by data analytics tools. The period is from 2018 until 2021. This study measures audit efficiency expressed in hours saved per data analytics tool. In addition, the study analyses if the audit efficiency increases when multiple tools are used by one client or a client uses one tool for multiple times. In total there are 606 useful observations and 22 different data analytics tools used during these three years. I categorize the 22 data analytics tools into 20 due to the number of observations.

I add an independent variable to my data named *SIZE*. I choose *SIZE* as an independent variable, because I expect that the size of the firm may affect the impact on audit efficiency. Since larger firms usually have more resources, more complex operations and more procedures for which tools can be used. In the audit practice it is more common that multiple tools are used and could be applied for the audit of medium and large sized firms. I classify all clients on firm sizes, based on the financial information accessible for auditors on *company.info*. In total four size categories are applicable for firms, namely Micro, Small, Medium and Large. I also add the category Not determined for clients for whom I may not be able to determine the size. Finally, I aggregate the categories Micro and Small, which leads to four categories of *SIZE* in total.

Besides the impact of the size of the client, I also analyze for the impact of the *INDUSTRY* on the audit efficiency. I categorize my industries from 26 into two industries, which is in line with the categorization applied by the Big Four company I work for. These two categories are Generic and Financial Services. I expect that for the audit of financial services clients, also data analytics tools are applied that are more useful for this industry, like for specific procedures regarding fair value accounting or other complex models. I therefore expect that the type of industry is decisive for the type of tool(s), which subsequently determines the degree of impact on audit efficiency.

Last, I perform an additional analysis as a follow up on the main analyses by analyzing the impact of the independent variable *AUDITIMPACT* on audit efficiency. The data contains in total seven types of audit impact. For testing purposes, I categorize the seven audit impact assessments from seven to four. This because grouping categories based on their nature seems more logical for the results of the tests.

I first perform a one sample *t*-test and run a bootstrap in order to measure the impact of data analytics on audit efficiency, which is expressed in hours saved. A bootstrap ensures that statistical values and analytical models are reliable and give accurate results (DASC, 2022).

Second, I analyze by a one-way ANOVA test whether the impact of data analytics on audit efficiency depends on data analytics tools. I test if the results significantly differ from zero, suggesting that data analytics positively impact audit efficiency. Additionally, I analyze whether the use of multiple tools by one client or the use of one tool for multiple times increases audit efficiency. Subsequently, I perform a one-way ANOVA test for the impact of the variable *SIZE* on audit efficiency. I am also performing a Pairwise comparison to analyze which firm sizes significantly differ from each other. Thirdly, I perform a one-way ANOVA test on the variable *INDUSTRY*, to test whether the type of industry significantly impact audit efficiency. The last one-way ANOVA test is to analyze additionally if the variable audit impact effects audit efficiency.

Overall, a mean value of 17.38 is found for *HOURSSAVED* and this value significantly deviates from zero. For the use of multiple tools or one tool for multiple times, I find that with a mean value of 20.84 multiple use of a tool or use of multiple tools significantly differs from one tool or use of a tool once (M = 12.58). These findings suggest that that data analytics positively impacts audit efficiency and even increases audit efficiency using multiple tools or one tool for multiple times, measured in hours saved by the tool. I find for the factor *SIZE* that the effect on the total variation in audit efficiency is significant. Given the results of the Pairwise Comparison test, I find that Medium and Large sized firms significantly differ. These two categories contain the most observations of the total of 606. For the variable *INDUSTRY*, the study finds that the effect on the total variation in audit efficiency is significant. As there are only two categories for industry, a pairwise comparison test is not possible. Finally, the results of the variable *AUDITIMPACT* show that the category Not determined significantly differs with the other three categories. Overall, the effect on the total variation in audit efficiency is significant for this variable as well.

Prior studies examined this topic by conducting interviews and surveys due to a lack of information and data (Earley, 2015, p. 494); (Tiberius & Hirth, 2019, pp. 1-2), but there is a lack of archival evidence on this topic. One of the reasons for this is as mentioned a lack of information in public available data. The study of Ratz examines how data analytics can create value in audit procedures by means of a case study of a Big Four company in Luxembourg (Ratz, 2017). Ratz compares two data analytics audits of the same client in the first and second year of data analytics implementation. However, the study lacks in measuring the efficiency and quality of the audit. This is the part where my thesis contributes to the literature. Another prior study of Tiberius and Hirth (2019) finds that digitalization could save time for auditors and that auditors will be used for new and more complex tasks that cannot be automated. I find that indeed digitalization, in particular data analytics tools are time saving as it increases the audit efficiency in hours saved. Last, the study of Gao et al. examines the impact of data analytics on audit quality by using hand-collected information on data analytics skills of employees of audit firms via their LinkedIn profiles (Gao, Huang, & Wang, 2021). As there are similarities in findings, this study examines the impact on audit quality, while my thesis examines the impact on audit efficiency based on internally obtained data as a result of the implementation and use of data analytics tools. In addition, the effect of factors *SIZE, INDUSTRY* and *AUDITIMPACT* is also considered during the analysis.

This study therefore investigates the impact of data analytics within the audit by analyzing the data within a Big Four company. The focus area is the impact of data analytics on the audit efficiency. The investigation in data analytics is of the interest of every stakeholder of an audit firm. You might think of the shareholders, employees and suppliers. Also, students and professors who would have to adapt to this in terms of education. Next the audit clients, investors and regulators will experience if data analytics will (partially) replace auditors work or how this could affect their businesses. Audit clients should also invest in data analytics to properly prepare their administration or to reduce the tasks of the auditor. These changes need to be in line with the regulations and standards (IASB). The regulators need insights on regulatory reforms, to adapt to the technological changes and adjust their supervision on audit firms and audit clients. For investors a new investment opportunity exists and so on.

The main contribution of this thesis is to provide quantitative perspective to this debate that is mainly based on qualitative research. This thesis can provide reporting organizations, users of reports, auditors and Information Technology (IT) auditors, investors, university professors, students and regulators with valuable information regarding the investigation in data analytics and compliance with laws and regulations. If procedures of the auditor could be computerized, audit firms need to invest in data analytical solutions and hire and/or educate staff to apply these data analytical solutions. If auditors are lagging behind the digital revolution, this could lead to reputational damage and delay in technological development (El Messaoudi & Kintou, 2019).

This study has its limitations. The data is obtained from one Big Four company and therefore has its limitation in variety of audit firms. The sample size used in this study is relatively small, because the obtained data consist of in total 6.520 observations and I only include 606 observations. This due to the lack of hours saved by data analytics tools for the remaining population. Another limitation is that I only collect data regarding the kind of data analytics tools developed and implemented by one company. I further analyze only three years of data which also limits my sample size. I also lack in analyzing the differences in audit efficiency between the years, so from first implementation year to the last observed year. This is because data analytics tools are not implemented for all and mainly the same clients starting from 2018 to compare the use and efficiency for a time period of two or three years for example. Also, the observations are not equally divided over the entire population to make and conclude over the comparisons. The study also does not have data regarding the audit efficiency before implementation of data analytics tools to compare the hours saved with the period before implementation. Future research could extend the sample period and obtain more data per client to examine the differences per client for more than two years at least.

This study consists of eight chapters in total. The first chapter is the introduction of this thesis, as described above. Chapter 2 explains the key literature. The key literature is subdivided into three sections. Chapter 2.1 until 2.3 explains concepts and evidence of prior literature. Based on these chapters, chapter 3 describes the hypotheses development. Chapter 4 explains the research setting. Then chapter 5 explains the methodology, the research design, the theoretical constructs, which model this research uses and the collection of data. Chapter 6 answers the main research question based on the results, the analysis and interpretation of the results and the findings. Chapter 7 investigates an additional analysis. Finally, chapter 8 concludes the main results, limitations and recommendations for further research.

2. Key Literature

This section discusses existing literature on relevant topics for this research. As there is a gap in literature regarding the impact of data analytics on audit efficiency and the relevance for the audit profession, this topic is further investigated. The thesis uses mainly four articles as a starting point. Three of the papers examine data analytics in auditing and one research examines audit efficiency. Only one paper examines the impact of data analytics in through an empirical research. Therefore, this research considers these papers and especially the last paper as the starting point and as the literature review. This research examines the impact of data analytics on the efficiency of the audit. The key literature is divided into three parts. The first two paragraphs illustrate the theory and concepts of data analytics and audit efficiency. The third paragraph examines existing literature.

2.1 Data analytics (in auditing)

Before discussing the role of data analytics in auditing, first it is important to define the term data analytics. According to Gartner (2021), data analytics "is the management of data for all operational and analytical uses and the analysis of data to drive business processes. It is applied to improve business outcomes through more effective decision making and enhances customer experiences". Often, data analytics is discussed in conjunction with the concept big data (Earley, 2015). Data analytics and big data are two independent concepts, which can be interrelated (Alles & Gray, 2016). Big data is defined as "high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhances insight, decision making and process automation" (Gartner, 2021). Volume, velocity and variety are often referred to as the three V's of big data (Earley, 2015). According to Rajaraman, two more characteristics define big data, which are veracity and value (Rajaraman, 2016). The five V's are defined by Rajaraman as follows:

1. Volume: Amount of global digital data created, replicated and consumed.
2. Variety: Refers to text, numbers, images, video data and audio. Also, the broad scope of data (Alles & Gray, 2016).
3. Velocity: The speed of or frequency of change in data and its relevance.
4. Veracity: Accuracy and integrity of data.
5. Value: Data gains value when it provides information for further action (Rajaraman, 2016).

According to Rajamaran (2016), "data analytics is concerned with extraction of actionable knowledge and insights from big data". Rajaraman describes four types of data analytics.

1. Descriptive analytics: Data gathered and presented as a chart or image to provide insight into the content of the data. It reflects information in the past.
2. Predictive analytics: By extrapolating available data, a prediction is made for the future.
3. Exploratory or discovery analytics: Collecting data and analyzing these data provides insights, accidental discoveries and pattern discoveries.
4. Prescriptive analytics: Collected data offers opportunities to optimize solutions for existing problems (Rajaraman, 2016).

Using technological tools and software programs, patterns or relations are identified among data (Earley, 2015). This is where the expertise of auditors is required namely, to interpret the output and results of the analysis performed by the tool.

With the use of data analytics tools, auditors can test 100 percent of the population instead of sampling the population (Earley, 2015). This enables the auditors to help assess risks and identify trends through the audit process. The most common analytics applied in the audit profession are descriptive analytics (Murray, 2017). The results can be used to produce red flags indicating potential control breakdowns or to arrange populations or other data. Diagnostic analytics or discovery analytics can give reasons behind the findings of descriptive analytics (Murray, 2017). Predictive and prescriptive analytics are more used in suggesting potential future outcomes and respectively help humans in making better decisions. As descriptive and diagnostic analytics provide insights and reasons behind the findings, based on gathered data, these tools are the most interesting for the audit practice.

2.2. Efficiency

The definition of efficiency is achieving something with the least waste of time and effort (Dictionary.com, 2012). The IAASB argues that the way of work performed in practice can be an important factor in both effectiveness and efficiency (International Auditing and Assurance Standards Board, 2014).

A practice guide by The Institute of Internal Auditors provides guidance on measuring the effectiveness and efficiency to internal audit activities (Dubis, Jain, Manchanda, & Thakkar, 2010). The next paragraph defines the performance measures to measure efficiency according to Dubis et al. (2010):

- “Number of audits scheduled.
- Number of audits completed.
- Timeliness of performance feedback.

- Staff utilization – direct vs. indirect time.
- Completed audits per auditor.
- Actual hours vs. Budgeted hours.
- Audit report cycle time: elapsed time from opening conference to fieldwork completion and elapse time from fieldwork completion to final report.
- Number of internal audit reports issued vs. planned internal audits".

Further measures of variables of efficiency are explained in this paragraph. The study of Ludwig examines the impact of positive surprise in audit efficiency and defines the measurement of variables of efficiency (Ludwig, 2000). Positive surprises should result in a reduction in audit testing. The study investigates that reduction in the extent of audit testing corresponds to audit efficiency. Measures of efficiency are sample size and detection of positive surprise. Sample size is the most appropriate measure of extent of testing. The extent of testing may be associated with the time required to complete the task, which is also in line with the definition of efficiency as mentioned in this chapter. Time to complete an audit has been a frequently used measure for audit efficiency (Ludwig, 2000).

2.3. Literature review

There are several practitioner articles researching the future of data analytics in auditing (Earley, 2015). An example is a survey by KPMG from 2014 until 2015 in 15 countries, in which they interview 830 senior business executives regarding their thoughts about data analytics (KPMG, 2015)[1]. Also, another big four company PwC examined the use of data analytics concluding that students must learn to become data scientists (PwC, 2015)[2]. Nowadays we also see in the education for chartered accountant that information technology is playing an increasingly important role. The expectation is that organizations will lag its competitors if they do not invest in a timely manner in data analytics (Earley, 2015). However, there is limited empirical academic research regarding this topic (Earley, 2015).

Ratz (2017) examines how data analytics can create value in audit procedures by means of a case study of a Big 4 four firm in Luxembourg. The author compares two data analytics audits of the same client and use the prior year as the learning experience and action initiator for the upcoming audit. The study improves the collaboration of IT auditors and financial auditors, which is called the

[1] KPMG is a Swiss entity with which the independent member firms of the KPMG network are affiliated (KPMG, 2015). The entity is one of the Big Four companies and provides services regarding Tax, Audit and Advisory.

[2] PwC is a Big four company providing audit and assurance, consulting and tax services (PwC, 2015).

shared mental model. This is because auditors are lacking in knowledge about information technology (IT). These different approaches allowed the researcher in gathering both qualitative and quantitative data.

The study observed that during the first-year data analytics was implemented, the quality increased, while time and budget were exceeded (Ratz, 2017). This is also argued by the International Auditing and Assurance Standards Board (IAASB) that these techniques can be time consuming in the first year (International Auditing and Assurance Standards Board, 2014). Based on the prior year audit an action plan is performed by the engagement team. The action plan is drawn up based on problems arise, feedback on prior year, survey on financial and IT auditors, clients' understanding and opinion and based on literature. Next the action plan is implemented to incorporate a solution for the current year audit. In the second-year quality increased even more and time and budget were as expected. After the audit the action plan is evaluated to identify the added value of data analytics in the current year compared to prior year.

The study concludes that auditor's review and interpretation are essential in the audit, because machines can also make mistakes. In addition, a common understanding of IT and financial auditors is crucial to create added value to the audit. Therefore, the auditor will not be replaced in the short term by technological tools (Ratz, 2017). Overall, this case study suggests that data analytics could add value by increasing efficiency and quality of the audit.

Tiberius and Hirth (2019) use a survey among German auditing professionals to examine expected changes in auditing practices due to digitalization within the next five to ten years. In this study, they ask auditors, university professors of auditing, regulators and IT audit software experts to evaluate twenty projections about digitization (Tiberius & Hirth, 2019). The method applied by the researchers is the Delphi method. The Delphi method is a research method that seeks the opinion of experts on a subject on which there is no consensus over at least two rounds. The study uses a standardized questionnaire and by means of structured feedback about the results from the first-round respondents', consensus is enhanced. The German auditing market is the aim of the study and this method determines the impact of digitization on auditing by generating a future scenario. By merging multiple expert statements, the most probable scenario is generated.

As stated before, the questionnaire contains twenty projections. With a four-point Likert scale the results are obtained by asking respondents whether to agree or disagree with the statements. Based on the results, most of the expert group believes that technological progress will not occur very quickly and therefore does not see massive job losses for auditors (Tiberius & Hirth, 2019). Thus, it will not threaten the role of humans in the auditing process. Real-time corporate reporting will be the new standard as a result of technological progress. Continuous auditing instead of annual auditing will be the approach for the future. Still, personal judgment of auditors will be necessary for the manual procedures to rely on. Also, decision making, or consulting are tasks that cannot be automated. Therefore, auditing procedures which are automated could be time saving for auditors, but auditors will be still needed for new and more complex tasks that cannot be automated. The researchers' conclusion is that new technologies such as data analytics are unlikely to fully replace audit work but will provide relief and support (Tiberius & Hirth, 2019).

The third study examines the contribution of a surprise between audit planning and substantive testing to audit efficiency (Ludwig, 2000). Ludwig develops a model of factors that signals an opportunity to reduce audit procedures. The study uses a field experiment to gather participant data to assess the importance of environmental and individual auditor variables associated with audit efficiency. Herewith the impact of supervisor behavior and analytical procedures are used. The researcher prepared 69 sets of case materials which were distributed to four regional public accounting firms. All materials were mailed to the participants whereof 54 participants provided complete data and 77% of these responses were useful. The results of the experiment were that supervisor behavior directly relates to audit efficiency and analytical procedures indirectly relate to audit efficiency. There is a relation between participants who did perform the analytical procedure as they were more likely to consult with the supervisor regarding the surprise, than those who did not perform analytical procedures. Therefore, analytical procedures indirectly influence audit efficiency.

The last study examines the impact of data analytics on audit quality by using hand-collected information on data analytics skills of employees of audit firms (Gao, Huang, & Wang, 2021). The authors measure the data analytics capability of audit offices by gathering the number of employees who disclose relevant data analytics skills on their LinkedIn profiles. The measure is based on skills disclosed in December 2015. To capture audit quality, the authors use accounting restatements of clients because misstatements are directly under the auditor's influence. Audit client engagement and accounting restatement data is from 2011 to 2015. The study constitutes five tests.

First, the study tests the relation between data analytics and audit quality by group comparisons between time period, industries and audit offices. Second, the study tests whether the positive effect of data analytics on audit quality is driven by frontline auditors or non-frontline employees. Next, the study tests the marginal benefit of data analytics on audit quality for clients with complex business operations. Fourth, the study examines the difficulty or uncertainty associated with accounting estimates that moderates the effect of data analytics. Last, the study tests whether the effect of data analytics is stronger for more digitalized clients.

The results of the study suggests that clients of audit firms with data analytics expertise issue fewer accounting restatements than client of audit firms with less expertise. Secondly, the findings suggests that the ability to apply data analytics by frontline auditors directly improves audit quality. The effect of data analytics support from non-frontline employees can indirectly improve the competency of frontline auditors. Thirdly, the results suggests that the marginal benefits of data analytics on audit quality are more substantial for clients with complex business operations. Fourthly, the tests suggests that the positive association between data analytics and audit quality is stronger in settings with difficult accounting estimates. Last, the study suggests that the positive association between data analytics and audit quality is stronger for client companies in digitalized industries (Gao, Huang, & Wang, 2021). The study finds that human capital investment in data analytics improves audit quality.

This thesis examines the impact of data analytics on the audit efficiency using more detailed data from one of the Big 4 audit firms to conduct an empirical study. The thesis therefore adds insights to the four studies I discuss above, which are based on survey data and one empirical research. Based on the available data, this research uses the variables of the data to measure audit efficiency. Using data, resulting from implemented data analytics techniques/tools in an audit firm, this thesis provides more accurate and reliable results compared to existing literature, which are primarily based on literature research, surveys, and case studies. Also, existing research does not examine the impact of data analytics tools on the audit efficiency. Therefore, this research provides new insights to the literature. Both in the field of the methodology and in the field of a new research question.

3. Hypothesis development

There is ongoing research regarding the use of data analytics in the audit practice. Data analytics is aimed at analyzing data with the aim of gaining insight into the processes to which that data relates. It is a process of examining, cleaning and using the data to analyze, interpret and document the valuable information. This is made possible by using technological tools and software programs to identify patterns or relations or for sampling purposes for example (Earley, 2015). Prior studies examine the relation between data analytics in the audit profession.

The study of Ratz examines the use of data analytics for a client audited by a Big four firm for two audit years. Within this study a comparison for the same client between two years is made. The results suggest that data analytics increase the quality of the audit in both years (Ratz, 2017). Also, a common understanding of IT and financial auditors is crucial to create added value to the audit. These findings are strengthened by the empirical research of Gao, Huang and Wang. The study suggests that data analytics support of non-frontline employees improves the competency of frontline auditors. Both frontline auditors and non-frontline employees' knowledge and use of data analytics positively impacts audit quality (Gao, Huang, & Wang, 2021). The increase in audit quality is measured through various tests resulting in, fewer accounting restatements with data analytics expertise, marginal benefits of data analytics on audit quality for complex business operations, positive association between data analytics and audit quality for difficult accounting estimates and for clients in digitalized industries. The study of Gao et al. finds that data analytics improves audit quality by human capital investment. These two studies support that data analytics positively impacts audit quality.

As data analytics increase the audit quality, I expect that data analytics has a positive impact on audit efficiency. Data analytics is applied to improve business outcomes and enhances customer experiences (Gartner, 2021). I believe that based on experience gained over the years, data analytics is time saving and provides more insights. Recent studies support that data analytics could save time for auditors (Tiberius & Hirth, 2019) and that it reduces the extent of audit testing (Ludwig, 2000). As both are measures of audit efficiency, these existing studies strengthen the expectation of this study that the use of data analytics improves audit efficiency. As there is little and no empirical evidence on which to base expectations, I formulate the hypothesis of this study in the alternative form.

Hypothesis 1: Data analytics has a positive impact on audit efficiency.

The first hypothesis quantitatively tests the relation between data analytics and audit efficiency. It tests whether the mean of *HOURSSAVED* per data analytics tool is significantly different from zero. Subsequently, I examine whether the impact of data analytics depends on the independent variable size. The reason for choosing *SIZE* as an independent variable is based on prior research. The research of Gao et al. demonstrates that the effect of data analytics on audit quality is stronger for audit clients with more complex operations. The size of the firm may affect the impact on audit efficiency since larger firms usually have more resources, more complex operations and more procedures for which tools can be used. The firm size is determined based on the following three criteria, revenue, total assets and total employees. These three criteria are the size criteria applied for the financial statements as set by Article 397 Burgerlijk Wetboek 2 lid 2 (Braam, 2016).

I categorize the clients based on firm sizes, to identify any differences between firm sizes. I expect that most of the tools are implemented for medium and large sized firms, in contrast to micro and small firms. Because, in practice it is more common that (more) tools are used and could be applied for the audit of medium and large sized firms. Based on the definition of *SIZE* of the firm in the audit and based on my experience within the audit, I can assume that *SIZE* is associated with the complexity of a client. Therefore, expecting that *SIZE* of the client has an impact on the results is in line with Gao et al. Therefore, I formulate my second hypothesis in the alternative form as follows:

Hypothesis 2: The impact of data analytics on audit efficiency depends on client size.

I quantitatively test for my second hypothesis the relation between size and audit efficiency. Testing for the second hypothesis should provide better insights of data analytics tools on the audit efficiency effected by the size of the firm.

Last, I examine whether the impact of data analytics depends on the independent variable industry. I choose *INDUSTRY* as an independent variable, because I expect that the type of industry is decisive for the type of tool(s), which subsequently determines the degree of impact on audit efficiency. The variable *INDUSTRY* is categorized in Generic and Financial Services (FS). The generic industry contains all industries except financial services. Financial services consist of mainly banks, insurers, fund managers and pension funds (Deloitte, 2022). This categorization is applied within the audit practice and therefore applied by the company I work for. Financial services clients mainly rely on intricate computational models for fair value accounting for example[3]. I therefore expect that for the audit of financial services clients, also data analytics tools are applied that are more useful for

[3] The term 'fair value accounting' is: "the price that would be received to sell an asset or paid to transfer a liability in an orderly transaction between market participants at the measurement date (an exit price)" (IFRS, 2022). "When measuring fair value, an entity uses the assumptions that market participants would use when pricing the asset or the liability under current market conditions, including assumptions about risk" (IFRS, 2022).

this industry, like for procedures regarding fair value accounting or other complex models. I hereby expect that these industry specific data analytics tools positively impact audit efficiency.

This expectation is in line with the results of the study by Gao et al. The study suggests that the positive association between data analytics and audit quality is stronger in settings with difficult accounting estimates (Gao, Huang, & Wang, 2021). Especially for fair value items, because estimates rely on intricate computation models and therefore data analytics alleviates this estimation uncertainty. Gao et al. finds that the main results are stronger for clients with greater proportions of assets using fair value accounting. I expect a larger impact on audit efficiency due to data analytics tools applied in the audit of Financial Services clients. I also test the third hypothesis quantitatively. As such, I formulate the third hypothesis in the alternative form, as below:

Hypothesis 3: The impact of data analytics on audit efficiency differs by industry type.

4. Research setting

This chapter outlines the implementation of data analytics within one of the big four companies in the Netherlands, hereafter referred to as "Company X". Professionals have different opinions and views on the impact of data analytics on the audit profession. Some argue that data analytics will fully replace the auditor in the future (Moll & Yigitbasioglu, 2019). Others argue that data analytics cannot fully replace auditors. In order to outline the background and understand the use of data analytics within the audit profession, I analyze the implementation of data analytics within Company X.

Company X implemented the use of data analytics tools first in 2018. Due to the implementation, Company X set up a separate department within the Netherlands. Also, a department has been set up in India to support Dutch audit teams. The contact between the audit teams in the Netherlands and with India takes place via a digital platform. Company X uses data analytics tools in order to facilitate audit work. The tools are developed to for example match purchase orders, goods receipts and invoices. Besides that, another commonly used solution reconciles PDF documents, such as bank statements. To make use of these tools, you first need to submit via the channel of Company X to the tool you want to make use of. During the submission, you need to fill in for example the name of the engagement, the financial year of the client, the tool(s) you want to make use of and the expected hours to save due to the use of the tool(s). Finally, when you are finished with the use of the tool you fill in the actual hours saved or not saved by the tool(s) and you also assess the audit impact of the tool(s). The effectiveness and efficiency of the tool(s) are measured via the actual hours and audit impact submitted by the audit team. Refer to Chapter 7 for the additional analysis performed on audit impact.

To obtain the data regarding the submission and outcome of these tools, I first contact the 'Digital assurance and innovation specialist' within Company X. I set up a meeting with the specialist to analyze and extract the data that is useful for my thesis. I select the variables that are useful to test the efficiency of these tools. Refer to Chapter 5 for the details.

Besides the data to test for audit efficiency, I also obtain data to analyze the development in employment. Together with the specialist, I extract the employment data of the newly founded department in India to analyze the development. Prior studies examined that auditor's review, interpretation and personal judgement based on manual procedures are essential and reliable in the audit (Ratz, 2017) (Tiberius & Hirth, 2019). As there is no prior empirical evidence regarding this relation, this thesis expects that data analytics leads to an increase of the employment in the audit profession. The expectation is that more expertise is required in the field of data analytics, which leads to job creation. In the audit profession a decline in the employment is not expected, rather the audit profession is being educated for data analytics. I already experience during the education for the chartered accountant and on job level that there is more attention for data analytics and information technology in the past few years.

The data that I obtain with the specialist relates to the period from 2018 until 2021. In total 34 employees have been hired from 2018 until 2021. We can observe that the most employees have been hired in 2018 and 2020. The lowest employees have been hired in 2021. That 2018 is one of the years in which most of the employees are hired is logical. This because 2018 was the first year of implementing the data analytics tools in Company X. Besides that, in total 11 employees left Company X within these three years, whereof ten left in 2021. Still there are 23 employees working for this department.

Based on the data, I observe that data analytics positively impacts the employment, because a separate department has been set up in India. Besides that, a separate department has been set up in the Netherlands due to data analytics for the audit. I also experience as an auditor that in the last years auditors are educated more in the field of data analytics and IT. There is more attention and interest to data analytics and IT. As from these observations and data, I observe that data analytics leads to job creation within the audit function rather than reducing employment. Other studies find that technological progress will not occur very quickly, machines can make mistakes, so human involvement in the audit process is still essential. Also, complex tasks such as decision making, or consulting cannot be automated. Therefore, new technologies will provide relief and support, but are unlikely to fully replace the audit (Tiberius & Hirth, 2019). Summarizing that, based on prior research, data obtained and experience within the audit, I do not expect that data analytics will take over the role of the auditor. Given the under-staffing problems at audit companies, the use of data analytics has not led to negative employments effects so far. In contrary, we observe new job creations, support of auditor's tasks and education of auditors in the field of IT and data analytics.

5. Methodology

This chapter discusses the data and method applied to test the hypotheses as discussed in the previous chapter. I first discuss the data collection and sample selection, leading to 606 usable observations. Next, I discuss the variables used in the analyses. In the last section, I descriptively analyze the variables.

5.1 Data collection and sample selection

I first contact the 'Digital assurance and innovation specialist' in the department within the audit firm I work for, to gather data regarding data analytics tools used within the firm. The specialist has access to the data through the application they use. The application imports data routines of the data analytics tools used. As mentioned in the key literature, there is a gap in empirical research, that examines the impact of data analytics on the audit efficiency. This mainly relates to the shortage of data and variables in the existing databases to examine this relation.

I obtain the data from Company X in the Netherlands. Together with the specialist, I select the useful variables for my sample to further investigate. I select and extract the data starting from the year 2018 until 2021, as 2018 is the first year of implementation of the data analytics tools within Company X. In total there are 22 different tools used for clients of Company X over three audit years. The data consists of the following useful statistics: the financial year in which the tool is applied for the audit of the client, the name of the client, the industry in which the client operates, the tool used, the impact of the tool on the audit, the estimated hours and the saved hours. Empirically, I create the variable size by including the size of the firm to my data based on the following three criteria, revenue, total assets and total employees as set by Article 397 Burgerlijk Wetboek 2 lid 2 (Braam, 2016). I obtain these financial data manually via "company.info" as these are made available for auditors.

The total data consists of 6,520-line items. In my sample, I only include the line items where a value exists in the columns 'estimated hours saved' and 'hours saved'. Because I use hours saved as the primary measure of audit efficiency. Blank line items are excluded from my sample, as these do not add value to my sample to test for efficiency. This results in a total of 606 useful observations for 365 clients. The data is divided as follows: 2018 – 2019: 104, 2019 – 2020: 317 and 2020 – 2021: 185. The most observations relate to year 2019-2020. I use the data-by-data analytics tool for a particular client in a year. In addition, I analyze the efficiency on client level using a tool once or at least two times. For my hypothesis and for the main effect, I do not analyze the differences over three years, for the same clients and based on the same data analytics tool used. Because this reduces my sample to 34 clients and 87 observations over three years. Therefore, not further specifying my data is more suitable for the analyses. With the current specified data, I analyze the relation between data analytics tools and audit efficiency. Besides that, I analyze the efficiency based on engagement level. Per engagement I analyze whether the quantity of tool(s) used leads to efficiency.

5.2 Variable description

This study uses *HOURSSAVED* as the primary measure of audit efficiency. Prior studies use time to complete an audit (Ludwig, 2000) and actual hours defined by Dubis et al. as a measure of audit efficiency. Therefore, using hours saved per tool as the dependent variable is sufficient for this study. As described in the previous chapter, the sample data consists of the following relevant variables, year, client name, size firm, industry, data analytics tools, audit impact, estimated hours and saved hours. *SOLUTION* is the variable for data analytics tools used and I have 606 observations spread over 22 tools. These tools are applied during the audit of a client's engagement. Based on the hours saved per solution, I measure if the use of these tools is time saving. The saved hours have a value between -8 and 80. Besides the relation between tools and hours saved, I measure per client if the use of a tool for multiple years or the use of more than one tool is efficient.

Subsequently, I measure for the independent variables *SIZE* and *INDUSTRY* the main effects on *HOURSSAVED*. There are in total five categories of the variable *SIZE.* These are (1) Micro, (2) Small, (3) Medium, (4) Large and (5) Not determined. As mentioned in previous chapter, these categories are determined for each observation manually based on the criteria of Article 397 Burgerlijk Wetboek 2 lid 2 (Braam, 2016). I categorize the firms based on information published on *company.info* regarding total revenues, total assets and total employees. Based on the available information I categorize 532 observations into the first four categories. For 74 observations the disclosure of the information was not available or not sufficient to categorize into one of the four categories. Therefore, I categorize these 74 observations into the category 'Not determined'. Based on the observations per category, I group my categories from five into four for testing purposes. I merge the categories 'Micro' and 'Small' together into 'Micro/Small', because the observations are 15 and respectively 17. Table 7 shows the observations of variable *SIZE* per category.

The second independent variable is *INDUSTRY.* I expect that *INDUSTRY* is another relevant variable that effects the results of *HOURSSAVED* based on the *SOLUTION* used. The variable *INDUSTRY* contains 26 types of industries. For the (1) Generic industry the data contains 523 observations and for (2) FS these are 83 observations. The grouping of the 26 industries in these two categories are reported in Appendix A – Sample composition by Industry.

This study applies logarithmic transformation to the dependent variable *HOURSSAVED* due to high skewness levels. For the two independent variables, I create dummy variables because of the existence of multiple categories in these two independent variables. I do not include the variables financial year, client name and estimated hours as independent variables to my analysis. Financial year is not included as an independent variable, because not all 365 companies used the tools during three audit years. It is not sufficient to specify the data for only the firms that made use of the tools during two or three years. Because this minimizes the data to only 87 observations. Besides, this minimization will not provide the insights as I expect to obtain with the effects and interactions as described in previous paragraph.

I only use saved hours as the dependent variable and not the estimated hours, as these values are based on expectations instead of what is measured. I expect that saved hours by itself is a well indicator to measure the audit efficiency. I also do not include the variable client's name as an independent variable as I do not expect that the name will have any effect on the results. I perform later an additional analysis in Chapter 7 on the independent variable *AUDITIMPACT* to analyze the relation with *HOURSSAVED* for the 606 observations. As mentioned in the research design, the audit impact is submitted by the audit team after the tool has been used. Therefore, I additionally analyze whether the impact of data analytics on audit efficiency differs by the kind of audit impact. A list of all variable definitions is shown in Appendix A.

5.3 Analyses and model

5.3.1 One-way ANOVA

This study analyzes the impact of data analytics on audit efficiency. The data includes various data analytics tools used within Company X and the effect of the use of these tools expressed in hours. The first model to use is the one-way analysis of variance (ANOVA) (Laerd Statistics , 2018) which is suitable for the first part of the analysis. This model determines whether there are any statistical differences between two or more independent groups.

In the first analysis the study measures the effect of the data analytics tools, independent variable *SOLUTION,* on the dependent variable *HOURSSAVED.* The dependent variable is at interval level as these are measures in hours. The variable *SOLUTION* contains 22 different tools. For three of the 22 tools, there are only one observation per tool, therefore I group these three tools into one group named 'Other', resulting in 20 tools in total. With the one-way ANOVA it is possible to measure if the tools significantly impact the saved hours. In addition, it measures 20 independent observations for each tool. For comparing two groups, the one-way ANOVA is equal to performing an 'One-sample *t*-test'. Therefore, I also perform a one-sample *t*-test regarding variable *SOLUTION.*

This study performs the one-way ANOVA for the following variables, *SOLUTION* on *HOURSSAVED, SIZE* on *HOURSSAVED* and *INDUSTRY* on *HOURSSAVED.* I analyze with these tests if the variables individually impact the hours saved per tool. All independent variables are nominal values. The variable *SOLUTION* indicates the effect of data analytics on mean *HOURSSAVED,* and thus tests the efficiency, hypothesis 1. For variable *SOLUTION,* I also test whether the kind of tool impact audit efficiency. The variable *SIZE* indicates the effect of the size of the client on mean *HOURSSAVED,* and tests hypothesis 2. Last, the variable *INDUSTRY* indicates the effect of the industry in which the client operates on the mean of *HOURSSAVED,* and tests hypothesis 3. The corresponding model for the one-way ANOVA is as below:

$$HOURSSAVED = \mu + SOLUTION + SIZE + INDUSTRY + \varepsilon$$

5.3.2 Additional analysis

This study investigates also if there is a relation between *AUDITIMPACT* and *HOURSSAVED.* The chosen method for this analysis is the one-way ANOVA. There are seven types of audit impacts observed in my data as a result of the tool used. The expectation is that there is a relation between the kind of *AUDITIMPACT* submitted by the audit team and the *HOURSSAVED* due to the tool applied to the client's engagement. There are six categories for which the audit impact is determined for a total of 222 observations. The undetermined observations are in total 384, which I categorize as 'Not determined'. Also, for this variable I merge my categories into four. I categorize the following categories into 'Other' which are, completely change audit approach, first year audit and new evidence that was previously not possible. These three are categorized to one group, because in all three categories the evidence or the approach is new for the observed year. In addition, I grouped the following two, had not been introduced yet and no impact on the approach, into one category named as 'No impact'. Because, in case of these two there is no measurable impact on the approach or change impacting the audit. The final categories are (1) changes from sample to full population, (2) change audit approach, (3) no impact and (4) not determined'. Refer to Chapter 7 for the observations per category of *AUDITIMPACT.* For this factor, I expect that the Not determined category significantly differs. Because I expect that the category Not determined contains the observations in hours saved, which do not necessarily have a relation on the audit approach. So, the tools lead to an efficiency in hours, but there is no necessary relation with the change in audit impact.

6. Results

In this chapter, I perform analyses to test hypotheses 1, 2 and 3 as discussed in previous chapter. In the first paragraph, I analyze the impact of data analytics tools on audit efficiency with a one-way ANOVA. I second examine the impact of the variable size and subsequently the impact of the industry on audit efficiency separately in two one-way ANOVA's. The results of the analyses are listed in the Appendix.

6.1 *H1: The impact of data analytics on audit efficiency*

To examine the impact of data analytics on audit efficiency, I analyze the impact of (the kind of) data analytics tools used by Company X from 2018 until 2021. I first perform my analysis with the independent variable *SOLUTION* to answer the main question if data analytics positively impacts audit efficiency. I expect that data analytics positively impacts audit efficiency.

In order to examine whether tools in general are associated with an increase in efficiency, I perform a one sample *t*-test to test the null hypothesis. Overall a mean value of 17.38 is found for *HOURSSAVED* and this value significantly deviates from zero, t (605) = 13.44, p < .001, Cohen's d = 31.84. Refer to Table 1 and 2 for the results. However, since this variable was extremely positively skewed (Z_{skewness} = 49.59), I also perform a bootstraps (10,000 bootstraps) estimation for the 95 percent confidence interval for the true population mean, CI_{95} = [14.94, 20.03] refer to Table 3.

The one-way ANOVA test with *SOLUTION* as independent variable and audit efficiency *HOURSSAVED* as dependent variable, is done in order to evaluate possible differences in means. Overall, differences in data analytics tools explains 18.3% of the total variation *HOURSSAVED* and this effect is significant, F (19, 586) = 6.91, p < .001, $\eta^2 = .18$, see Table 4. From this finding we can conclude that at least two data analytics tools differ from each other in means for *HOURSSAVED* (some means differ). Due to high skewness, I also perform a test with the transformed dependent variable HOURSSAVED. Table 5 shows the results, which are equal to the significance level obtained without logarithmic transformation. Table 6 shows the means per data analytics tool. Due to small sample sizes per condition not all means are reliable estimates for the population means. However, it shows for example that tool SIAC has the highest estimated mean value (M = 79.00, SD = 86.27) and tool DRB the lowest mean (M = 5.43, SD = 8.19). Overall, the results suggest that the kind of data analytics tools positively impact the level of audit efficiency.

In addition, I analyze at client level whether the use of a tool for multiple times or the use of multiple tools by one client increases the audit efficiency measured in *HOURSSAVED*. This results in a higher efficiency level for clients using multiple tools or one tool for multiple times with an average of 20.84 of *HOURSSAVED* in contrast to an average of 12.58 for clients using one tool or a tool for once. Differences in one or multiple tools used explains 1.6% of the total variation *HOURSSAVED* and this effect is significant, *F* (1, 604) = 10.09, *p* = .002, $\eta^2 = .02$, see Table 8 for the results. The results suggest that the increased use of data analytics tools significantly effects *HOURSSAVED*. All results of the tests suggest that the use of data analytics positively impacts and thus increases the audit efficiency. Hereby adding that making more use of data analytics also increases the audit efficiency.

6.2 *H2: The impact of size on audit efficiency*

The second one-way ANOVA is in order to test for hypothesis 2. The effect of the *SIZE* of the client explains about 2.0 % of total variation in *HOURSSAVED* and is significant, *F* (3, 602) = 4.07, *p* = .007, $\eta^2 = .02$, see Table 4. Post hoc test (Tukey's HSD) shows that only the mean *HOURSSAVED* for Medium size clients (*M* = 7.96, *SD* = 10.09) differs significantly from size Large (*M* = 20.20, *SD* = 36.17) with *p* = .004. Refer to Table 10 - Pairwise Comparisons. The other means are illustrated in Table 9. From this, the study can conclude that on average Medium sized clients score lower on *HOURSSAVED* than Large sized clients, but that other observed differences in *HOURSSAVED* means between different sizes should not be taken seriously. These results are in line with the expectation that most tools are used by Medium and Large sized firms, also considering the number of observations in both categories of size. Because it is more common in practice that the tools are used and could be used for these firm sizes.

6.3 *H3: The impact of industry on audit efficiency*

In this paragraph, I test regarding hypothesis 3 the effect of the *INDUSTRY* in which the client operates on audit efficiency. The effect *INDUSTRY* explains 1.0 % of total variation in *HOURSAVED* and is significant, *F* (1, 604) = 7.31, *p* = .007, $\eta^2 = .01$, see Table 4. For *INDUSTRY,* I cannot perform a post hoc test, as there are fewer than three groups. So, in general the effect is significant, but we cannot further compare which means significantly differ from each other. Additionally, I analyze whether specific tools used by Financial Services ensures a higher impact on audit efficiency. Table 12 shows the results of *HOURSSAVED* by industry. Only tool CM is more efficient for the Financial Services industry. All other tools are more efficient or almost equal in efficiency for the Generic industry.

7. Additional Analysis

In this section, I follow up on the main analyses of Chapter 6 and I perform an additional analysis on the relation between the independent variable *AUDITIMPACT* and *HOURSSAVED* as the dependent variable. This is a validation of the subjective audit impact by analyzing whether *HOURSSAVED* differs by different audit impact assessments. Also, for this analysis I perform a one-way ANOVA test in order to evaluate possible differences in means. The effect of *AUDITIMPACT* explains about 8.0 % of total variation in *HOURSSAVED* and is significant, F (3, 601) = 18.046, p = < .001, $\eta^2 = .08$, refer to model (1). Due to high skewness, I also perform a test with the transformed dependent variable HOURSSAVED. Model (1) shows the results, which are equal to the significance level obtained without logarithmic transformation.

(1) Results of One-way ANOVA's

	Dependent Variable: Hours Saved				
Factor	*F*	*df between*	*df within*	*p*	*η^2*
Audit Impact	18.046	3	601	< .001	.08
Audit Impact (transformed Hours Saved)	33.70	3	601	< .001	.14

Post hoc test (Tukey's HSD) shows that the mean *HOURSSAVED* for Not determined (*M* = 10.70, *SD* = 21.62) differs significantly from Changes from sample to full population (*M* = 27.02, *SD* = 22.17) with *p* = .002, see model (2). Not determined also significantly differs from Other (*M* = 41.29, *SD* = 58.80) with *p* = < .001 and No impact (*M* = 27.99, *SD* = 44.24) with *p* = < .001. Refer to model (2) for other means. From this, the study can conclude that on average Not determined audit impact scores lower on *HOURSSAVED* than Changes from sample to full population, Other and No impact as well, but that other observed differences in *HOURSSAVED* means between different audit impacts should not be taken seriously. These results are in line with the expectation that not for all data analytics tools the audit impact on the audit approach could be determined. Therefore, the difference of category Not determined with the other categories is as expected.

(2) One-way ANOVA descriptives - Audit Impact:

Audit Impact	*n*	*M*	*SD*	*Min*	*Max*
Not determined	384	10.70	21.62	-5.00	250.00
Changes from sample to full population	51	27.02	22.17	4.00	150.00
Other	21	41.29	58.80	-8.00	250.00
No impact	149	27.99	44.24	-4.00	336.00
Total	606	17.39	31.86	-8.00	336.00

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *AUDITIMPACT* as independent variable. The model contains data of four kind of audit impacts obtained from one of the Big Four companies in the Netherlands. *n* is the observations. *M* is the mean in *HOURSSAVED. SD* is the standard deviation in *HOURSSAVED.* Min and Max are the range of hours saved. Post hoc test (Tukey's HSD) shows the mean in *HOURSSAVED* for the categories. Category 'Other' contains the following categories, completely change audit approach (*n* = 10), first year audit (*n* = 2) and new evidence that was previously not possible (*n* = 10). The category 'No impact' contains the categories, had not been introduced yet (*n* = 104) and no impact on the approach (*n* = 45).

8. Conclusion

This thesis examines the impact of data analytics on audit efficiency. I analyze the impact of data analytics by testing the impact of data analytics tools applied within Company X on the hours saved using the tools. With this I measure via hours saved the impact on audit efficiency. I mainly used the methods one-way ANOVA to obtain the results for my hypotheses tested. The data is internally obtained within Company X. Overall, I find significant differences in the results of the one-way ANOVA's.

First, I find by performing a one sample *t*-test and by running a bootstrap that data analytics tools positively impact the hours saved per client. Overall, a mean value of 17.38 is found for *HOURSSAVED* and this value significantly deviates from zero. In addition, I find that clients using multiple tools or one tool for multiple times increase audit efficiency in contrast to clients using one tool or a tool for once. The mean is significant which is for multiple tools M = 20.84 of *HOURSSAVED* and for one tool M = 12.58. Both results of the one-way ANOVA tests suggest that data analytics positively impact audit efficiency, as data analytics is measured by the data analytics tool on audit efficiency.

Second, I find by one-way ANOVA tests that the impact of data analytics on audit efficiency depends on both the size and industry. For the factor *SIZE*, I find that Medium and Large sized firms significantly differ. It is more common in practice that most of the tools are used and could be used for these firm sizes. Considering also the number of observations in categories Medium and Large, these results are in line with the expectation of the study.

Within the factor *INDUSTRY* I do not find differences in categories, as there are only two categories Generic and FS to compare, therefore no mean differences noted. I also do not find any relation in specific tools used for the industry Financial services that increases the impact on audit efficiency. Only tool CM is more efficient for the Financial Services industry. All other tools are more efficient or almost equal in efficiency for the Generic industry.

I additionally performed an analysis on the factor *AUDITIMPACT*, where I found a significant difference for category Not determined with the other three possible categories. This is in line with the expectation of the study, because not for all tools an audit impact could be determined on the approach of the audit. We could observe that in the most cases (n = 384) the impact on the approach could not be determined.

All findings together answer the research question *"What is the impact of data analytics on audit efficiency"* and the sub-questions if data analytics tools positively impact audit efficiency and if the impact is affected by the size or the industry. Data analytics, measure by data analytics tools increases the audit efficiency in general and besides that size, industry and audit impact also positively impact the audit efficiency.

The contribution of this study is that it is the first study that quantitatively expresses the relation between data analytics and audit efficiency. Until now most studies in this field conduct qualitative analyses on the possible impact of data analytics, as these studies mainly lack on public available information and data.

Ratz (2017) examines how data analytics can create value in audit procedures by means of a case study of a Big Four company in Luxembourg, by comparing two data analytics audits of the same client for two years. The study observed that during the first year of implementation of data analytics quality increased, but time and budget were exceeded. In the second year the quality increased more, but time and budget were as expected. Ratz (2017) suggests that data analytics could add value by increasing efficiency and quality of the audit. My findings suggest that an increase in audit efficiency is impacted by data analytics and thereby contributes to the literature next to the contribution of the study by Ratz (2017).

The second study is a survey by Tiberius and Hirth (2019) which examines the expected changes in the audit due to digitalization. The results of the study show that digitalization could save time for auditors and that auditors will be used for new and more complex tasks that cannot be automated. The findings of Tiberius and Hirth (2019) are in line with the findings of this thesis. The results of my thesis support that indeed digitalization, in particular data analytics tools are time saving as it increases the audit efficiency in hours saved. Also, the observation outlined in the Research setting support that the implementation of data analytics tools created new vacancies, as a new department is set up in Company X to support the auditors in the use of data analytics.

Third, the study of Gao et al. 2021 examines the impact of data analytics on audit quality by using hand-collected information on data analytics skills of employees of audit firms via their LinkedIn profiles. They find that both frontline and non-frontline auditors improve audit quality by applying data analytics. This is in line with the findings of this study but then related to audit efficiency, as we also see an improvement due to data analytics by auditors and an improvement via non-auditors which support the auditors in and with the use of these tools. With the last, I refer to the Research setting regarding the impact and changes on employment. The results of Gao et al. 2021 also suggest that the marginal effects of data analytics on audit quality are more substantial for clients with complex business operations. Also, this is in line with the findings of this thesis regarding the increase in audit efficiency. This thesis finds that the impact of data analytics tools significantly differs for Medium and Large sized firms. Medium and Large sized firms mainly have more complex business operations and therefore applying more data analytics tools is more common.

Overall, these findings are insightful for auditors, regulators, investors, students, university professors, IT auditors and other users of financial statements. It can provide valuable information for investing purposes in data analytics, regarding the education of potential auditors and compliance with laws and regulations due to the use of data analytics tools. This thesis confirms the findings of Gao et al. (2021) that data analytics contribute to the audit, by measuring this in audit efficiency instead of audit quality what is examined by Gao et al.

Despite the findings of this study, it has its limitations. The sample size used in this study is relatively small and only limited to one Big Four company. Of the total data of 6.520, I only include 606 observations due to the lack of hours saved by data analytics tools for the remaining population. Besides that, I only collect data regarding the data analytics tools implemented by one Big Four company. I further analyze only three years of data which also limits my sample size. I also lack in analyzing the differences in audit efficiency between the years, so from first implementation year to the last observed year. This is because data analytics tools are not implemented for all and mainly the same clients starting from 2018 to compare the use and efficiency for a time period of two or three years for example. Also, the observations are not equally divided over the entire population to make and conclude over the comparisons. The study also does not have data regarding the audit efficiency before implementation of data analytics tools to compare the hours saved with the period before implementation. Future research could extend the sample period and obtain more data per client to examine the differences per client for more than two years at least. It could also address any alternative explanations that this thesis could not address.

In conclusion, I extent literature on data analytics measures by first examining audit efficiency instead of audit quality and then further broaden my contribution by examining the effects of size, industry and audit impact. Specifically, my research suggests that data analytics tools positively impact the audit efficiency, especially for Medium and Large sized industries. Besides that, the results mainly indicate that these tools positively impact the efficiency of Generic industries. Finally, the results confirm both the first, second and third hypotheses as the findings show significant differences for all hypotheses tested. The mutually and detailed significances are already mentioned before in this chapter.

9. Bibliography

Alles, M., & Gray, G. L. (2016). *Incorporating big data in audits: Identifying inhibitors and a research agenda to address those inhibitors.* Newark: Elsevier.

Braam, D. (2016, January 1). *Externe Verslaggeving Groottecriteria jaarrekening.* Retrieved from Externeverslaggeving.nl: https://externeverslaggeving.nl/jaarverslag/jaarrekening/groottecriteria-jaarrekening-2/

Bruins, R. (2018, July 2). De accountant moet zichzelf opnieuw uitvinden. *Executive Finance*, p. 4.

DASC. (2022). *IBM SPSS BOOTSTRAPPING*. Retrieved from DASC: https://www.dasc.nl/software/ibm-spss-statistics/ibm-spss-bootstrapping/

Deakin, E. B., & Granof, M. H. (1974). *Regression Analysis as a Means of Determining Audit Sample Size.* American Accounting Association.

Deloitte. (2022). *Financial Services Industry*. Retrieved from Deloitte: https://www2.deloitte.com/nl/nl/industries/financial-services.html

Dictionary.com. (2012). *Efficiency*. Retrieved from Dictionary.com: https://www.dictionary.com/browse/efficiency?

Dubis, G. S., Jain, P., Manchanda, A., & Thakkar, R. (2010). *Measuring internal audit effectiveness and efficiency.* Altamonte Springs Florida: The institute of internal auditors.

Earley, C. E. (2015). *Data analytics in auditing: Opportunities and challengs.* Providence: Elsevier Inc.

Eilifsen, A., Kinserdal, F., Messier, W. F., & McKee, T. E. (2020). *An exploratory study into the use of audit data analytics on audit engagements.* Bergen: Accounting Horizons.

El Messaoudi, O., & Kintou, A. (2019, January 21). Wij moeten veranderen; liever gisteren dan vandaag! *Accountant*, p. 4.

Frey, C. B., & Osborne, M. A. (2016). *The future of employment: How susceptible are jobsto computerisation?* Oxford: Elsevier Inc.

Gao, T., Huang, S., & Wang, R. (2021). *Data analytics and audit quality.* Singapore : School of Accountancy Research Center .

Gartner. (2021). *Gartner Glossary - Big Data* . Retrieved from www.gartner.com: https://www.gartner.com/en/information-technology/glossary/big-data

Gartner. (2021). *Gartner Glossary - Data And Analytics*. Retrieved from www.gartner.com: https://www.gartner.com/en/information-technology/glossary/data-and-analytics

IFRS. (2022). *IFRS 13 Fair Value Measurement*. Retrieved from IFRS: https://www.ifrs.org/issued-standards/list-of-standards/ifrs-13-fair-value-measurement/#:~:text=IFRS%2013%20defines%20fair%20value%20as%20the%20price%20that%20would,date%20(an%20exit%20price).

International Auditing and Assurance Standards Board. (2014). *A framework for audit quality. Key elements that create an environment for audit quality.* New York: International Federation of Accountants.

Katz, D. M. (2014, April 15). *Regulators Fear Big Data Threatens Audit Quality*. Retrieved from CFO.com: https://www.cfo.com/auditing/2014/04/regulators-fear-big-data-threatens-audit-quality/

KPMG. (2015). *Going beyond the data: turning data from insights into value.* London: KPMG.

Laerd Statistics . (2018). *One-way ANOVA in SPSS Statistics*. Retrieved from statistics.laerd: https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php

Ludwig, S. E. (2000). *Audit efficiency: The impact of positive surprise* . Arkansas: University of Arkansas.

Market business news. (2021). *What is efficiency? Definition and meaning*. Retrieved from Marketbusinessnews.com: https://marketbusinessnews.com/financial-glossary/efficiency-definition-meaning/

Moll, J., & Yigitbasioglu, O. (2019). *The role of internet-related technologies in shaping the work of accountants: New directions for accounting research.* Manchester: Elsevier.

Murray, M. (2017, July 18). Data analytics becomes reality. *Corporate compliance insights*, p. 4.

No, W. G., Lee, K. K., Huang, F., & Li, Q. (2019). *Multidimensional Audit Data Selection (MADS): A Framework for Using Data Analytics in the Audit Data Selection Process.* New York: American Accounting Association.

PwC. (2015). *Data driven: What students need to succeed in a rapidly changing business world.* United States : PwC.

Rajaraman, V. (2016). *Big Data Analytics.* Bangalore: Resonance.

Ratz, L. (2017). *Data Analytics in Audits: How can Data Analytics in auditing create value?* Turku: Turun Kauppakorkeakoulu: Turku School of Economics.

Tiberius, V., & Hirth, S. (2019). *Impacts of digitization on auditing: A Delphi study for Germany.* Potsdam: Elsevier Inc.

Appendix A

Variable definitions

Variable	Description
AUDITIMPACT	The impact on the audit as a result of the data analytics tool used. In total 7 different types of impact.
CLIENTNAME	Name of the client. Observed in total 365 different clients.
YEAR	The year of implementation. Data contains three years: 2018 – 2019, 2019 – 2020, 2020 – 2021.
ESTIMATED HOURS SAVED	Total expected hours to be saved by the data analytics tool. Range of hours are between 1 and 61.
HOURSSAVED	Total hours saved by the data analytics tool. Range of hours are between -8 and 80.
INDUSTRY	The industry in which the client operates. Data contains 26 different industries.
SIZE	Size if the client determined by total revenues, total assets and total employees. Contains 5 categories: micro, small, medium, large and not determined.
SOLUTION	Data analytics tool developed and implemented by Company X. Data contains 22 tools.

The table describes all variable definitions included in the data obtained from one of the Big Four companies in the Netherlands, named as Company X.

Sample composition by Industry

Industry	*n*	Classification into
Automotive	23	Generic
Aviation	5	Generic
Banking (Leasing)	28	Financial Services
Business Services	40	Generic
Chemicals	8	Generic
Consumer	1	Generic
Consumer products	14	Generic
Financial Services	8	Financial Services
Financial Services (Investment Management)	8	Financial Services
Financial Services (Insurance)	3	Financial Services
General Banking	12	Financial Services
General Manufacturing	72	Generic
Healthcare	1	Generic
IG&H (Infrastructure, Government & Health)	2	Generic
Industrial Products	29	Generic
Insurances	1	Financial Services
Life Insurance	15	Financial Services
Mining	2	Generic
Non-Life Insurance	8	Financial Services
Other	229	Generic
Real Estate Development	9	Generic
Real Estate Investment	28	Generic
Retail	26	Generic
Services	2	Generic
Wholesale	30	Generic
WoCo (housing corporations)	2	Generic

The table illustrates all 26 industries categorized by one of the Big Four companies in the Netherlands. *n* are the observations which are 606 in total. In column three, I classified the industries in two categories as this classification is applied in practice in the departments.

Appendix B

Table 1: One-sample t-Test

						95% Confidence Interval of the Difference	
	t	*df*	*One-Sided p*	*Two-Sided p*	*M*	*Lower*	*Upper*
HOURSSAVED	13.44	605	< .001	< .001	17.38	14.84	19.92

One-sample *t*-test is performed with *HOURSSAVED* as dependent variable and *SOLUTION* as independent variable. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The data is obtained from one of the Big Four companies in the Netherlands. *t* compares the means of two populations and examines whether the mean of the population is statistically different from the hypothesized value. *df* is the degrees of freedom, which refers to the maximum number of logically independent values. The one-sided and two-sided *p* state that the means in the two groups are unequal. *M* is the mean in *HOURSSAVED. SD* is the standard deviation in *HOURSSAVED.* Lower and Upper limit in hours saved.

Table 2: One-sample Effect Sizes

				95% Confidence Interval of the Difference	
		Standardizer	*Point Estimate*	*Lower*	*Upper*
HOURSSAVED	Cohen's d	31.84	.55	.46	.63

One-sample effect in Cohen's d of the one-sample t-test performed. *HOURSSAVED* is the dependent variable and measures for audit efficiency and *SOLUTION* represents data analytics tools as the independent variable. The data is obtained from one of the Big Four companies in the Netherlands. Cohen's d shows the effect size in the population with respect to the null hypothesis, that there is no relation between data analytics and audit efficiency.

Table 3: Bootstrap for One-sample Test

					95% Confidence Interval of the Difference	
	M	*Bias*	*SE*	*p (2-tailed)*	*Lower*	*Upper*
HOURSSAVED	17.38	-.003	1.30	< .001	14.94	20.02

A bootstrap is performed as part of the one-sample t-test. *HOURSSAVED* is the dependent variable and measures for audit efficiency and *SOLUTION* represents data analytics tools as the independent variable. The data is obtained from one of the Big Four companies in the Netherlands. A bootstrap ensures that statistical values and analytical models are reliable and give accurate results. *M* is the mean in *HOURSSAVED. Bias* relates to the proportion of bootstrap estimates that are less than the observed statistic. *SE* is the standard error. *p* state that the means in the two groups are significant. Lower and Upper limit in hours saved.

Table 4: Results of One-way ANOVA's

	Dependent Variable: Hours Saved				
Factor	*F*	*df between*	*df within*	*p*	*η^2*
Solution	6.91	19	586	< .001	.18
Size	4.07	3	602	.007	.02
Industry	7.31	1	604	.007	.01

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *SOLUTION, SIZE* and *INDUSTRY* as independent variables. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The table contains data from one of the Big Four companies in the Netherlands. *F is* the variance in *HOURSSAVED. df* is the degrees of freedom, which refers to the maximum number of logically independent values. *p* state that the means in the two groups are significant. *Partial eta-squared* is the indication of the extent to which the variance of *HOURSSAVED* is explained by the independent variable.

Table 5: Results of One-way ANOVA's Transformed

	Dependent Variable: Hours Saved Transformed				
Factor	*F*	*df* between	*df within*	*p*	*η^2*
Solution	11.95	19	586	< .001	.28
Size	6.31	3	602	< .001	.03
Industry	16.09	1	604	< .001	.03

One-way ANOVA is performed with logarithmic transformation of *HOURSSAVED* as dependent variable and *SOLUTION, SIZE* and *INDUSTRY* as independent variables. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The table contains data from one of the Big Four companies in the Netherlands. *F is* the variance in *HOURSSAVED. df* is the degrees of freedom, which refers to the maximum number of logically independent values. *p* state that the means in the two groups are significant. *Partial eta-squared* is the indication of the extent to which the variance of *HOURSSAVED* is explained by the independent variable.

Table 6: One-way ANOVA descriptives - Solution:

Data Analytics Tool	*n*	*M*	*SD*	*Min*	*Max*
3 WMC	17	28.82	19.56	-8.00	80.00
AT	9	20.44	17.29	0.00	60.00
AS	51	29.04	38.81	-4.00	250.00
ACR	44	39.32	42.50	11.00	250.00
CM	59	32.24	54.04	0.00	336.00
DGRB	17	43.71	64.71	-5.00	250.00
DPR	12	7.33	6.71	1.00	20.00
DRB	182	5.43	8.19	-2.00	64.00
DC	20	9.45	13.90	0.00	48.00
GM	4	52.50	57.37	0.00	120.00
ICR	3	24.00	13.86	16.00	40.00
KP	9	18.56	26.54	0.00	80.00
KC	117	8.10	11.44	0.00	80.00
LSLG	8	23.00	19.68	3.00	64.00
LA	12	8.00	8.83	0.00	32.00
PDE	27	28.85	41.36	0.00	200.00
RCP	8	8.38	5.50	1.00	16.00
SB	2	16.00	0.00	16.00	16.00
SIAC	2	79.00	86.27	18.00	140.00
Other	3	7.00	1.73	5.00	8.00
Total	606	17.38	31.84	-8.00	336.00

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *SOLUTION* as independent variable. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The table contains data of 20 different data analytics tools applied in one of the Big Four companies in the Netherlands. *n* is the observations. *M* is the mean in *HOURSSAVED. SD* is the standard deviation in *HOURSSAVED.* Min and Max are the range of hours saved.

Table 7: One-way ANOVA descriptives – One or multiple tools:

Client using one or multiple tools	*n*	*M*	*SD*	*Min*	*Max*
Clients using 1 tool	254	12.58	20.80	-8.00	150.00
Clients using tool(s) for multiple times	352	20.84	37.50	-5.00	336.00
Total	606	17.38	31.84	-8.00	336.00

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *SOLUTION* as independent variable. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The table contains data obtained from one of the Big Four companies in the Netherlands. *n* is the observations. *M* is the mean in *HOURSSAVED. SD* is the standard deviation in *HOURSSAVED.* Min and Max are the range of hours saved. This ANOVA test illustrates the results of clients using one tool or a tool once in contrast to clients using multiple tools or one tools for multiple times.

Table 8: Results of One-way ANOVA – One or multiple tools

	Dependent Variable: Hours Saved				
Factor	*F*	*df between*	*df within*	*p*	*η^2*
Solution by one or multiple tools used	10.09	1	604	.002	.02

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *SOLUTION* as independent variable. *HOURSSAVED* measures for audit efficiency and *SOLUTION* represents data analytics tools. The table contains data obtained from one of the Big Four companies in the Netherlands. *n* is the observations. *F is* the variance in *HOURSSAVED. df* is the degrees of freedom, which refers to the maximum number of logically independent values. *p* state that the means in the two groups are significant. *Partial eta-squared* is the indication of the extent to which the variance of *HOURSSAVED* is explained by the independent variable. This ANOVA test illustrates the results of clients using one tool or a tool once in contrast to clients using multiple tools or one tools for multiple times.

Table 9: One-way ANOVA descriptives - Size firm:

Size	*n*	*M*	*SD*	*Min*	*Max*
Micro/Small	32	14.98	21.53	0.00	64.00
Medium	93	7.96	10.09	-0.00	64.00
Large	407	20.20	36.17	-5.00	336.00
Not determined	74	14.73	25.11	0.00	150.00
Total	606	17.38	31.84	-8.00	336.00

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *SIZE* as the independent variable. *HOURSSAVED* measures for audit efficiency. *n* is the observations*. M* is the mean in *HOURSSAVED*. *SD* is the standard deviation in *HOURSSAVED.* Min and Max are the range of hours saved. Post hoc test (Tukey's HSD) shows the mean in *HOURSSAVED* for the categories. The table contains data of five types of firm sizes from one of the Big Four companies in the Netherlands. Category Micro/Small contains 15 observations of Micro and 17 of Small. These two categories are aggregated into one category given the limited number of each individual category containing in total 32 observations. Category Not determined contains observations for which size cannot be determined as the information is not available on *company.info*. To determine the size of the firm, information on *company.info* has been consulted. Based on the size criteria set by Article 397 Burgerlijk Wetboek 2 lid 2 (Braam, 2016), total revenues, total assets and total employees, the size of the firm is determined.

Micro: Assets ≤ € 350k, Revenues ≤ € 700k and employees < 10.
Small: Assets ≤ € 6 mio, Revenues ≤ € 12 mio and employees < 50.
Medium: Assets ≤ € 20 mio, Revenues ≤ € 40 mio and employees < 250.
Large: Assets > 20 mio, Revenues > 40 mio and employees ≥ 250.

Table 10: Pairwise Comparisons - Size

		Dependent Variable: Hours Saved		
(I) Size	(J) Size	*M (I-J)*	*SE*	*p*
Micro/Small	Medium	7.03	6.48	.70
	Large	-5.21	5.80	.81
	Not determined	.25	6.68	1.00
Medium	Micro/Small	-7.03	6.48	.70
	Large	-12.24	3.63	.004*
	Not determined	-6.77	4.92	.52
Large	Micro/Small	5.21	5.80	.81
	Medium	12.24	3.63	.004*
	Not determined	5.47	3.99	.52
Not determined	Micro/Small	-.25	6.68	1.00
	Medium	6.77	4.92	.52
	Large	-5.47	3.99	.52

This table illustrates the pairwise comparisons as part of the one-way ANOVA test. *HOURSSAVED* is the dependent variable and measures for audit efficiency and *SIZE* is the independent variable. The table contains data obtained from one of the Big Four companies in the Netherlands. *M* are the mean differences in *HOURSSAVED* between the different firm sizes. *SE* is the standard error. *p* state that the means in the two groups are significant. *Size Medium significantly differs from size Large, $p < .004$.

Table 11: One-way ANOVA descriptives - Industry:

Industry	*n*	*M*	*SD*	*Min*	*Max*
Generic	523	15.99	29.46	-8.00	250.00
Financial Services (FS)	83	26.11	43.19	-5.00	336.00
Total	606	17.38	31.84	-8.00	336.00

One-way ANOVA is performed with *HOURSSAVED* as dependent variable and *INDUSTRY* as the independent variable. *HOURSSAVED* measures for audit efficiency. *n* is the observations. *M* is the mean in *HOURSSAVED*. *SD* is the standard deviation in *HOURSSAVED.* Min and Max are the range of hours saved. The table contains the data of 2 industries in one of the Big Four companies in the Netherlands.

Table 12: Total hours saved by industry by tool

Industry	Data Analytics Tool	*n*	*Total Hours Saved*
FS	3 WMC	1	24
	AT	4	58
	AS	8	203
	ACR	25	530
	CM	18	952
	DGRB	5	112
	DRB	11	76
	ICR	1	16
	KC	4	88
	LSLG	4	85
	SIAC	1	18
	Other	1	5
Generic	3 WMC	17	466
	AT	5	126
	AS	43	1278
	ACR	19	1200
	CM	41	950
	DGRB	12	631
	DPR	12	88
	DRB	171	913
	DC	20	189
	GM	4	210
	ICR	1	56
	KP	9	167
	KC	113	860
	LSLG	4	99
	LA	12	96
	PDE	27	779
	RCP	8	67
	SB	2	32
	SIAC	1	140
	Other	2	16

This table illustrates the total hours saved by each industry by tool. Overall, the table shows that industry Generic contains the most observations and the most in total hours saved.