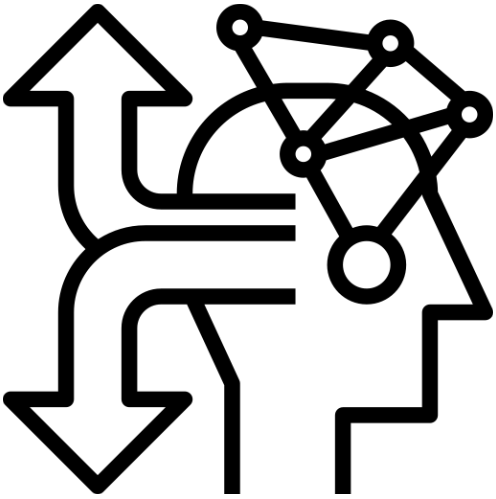


COGNITIVE LOAD IN DATA-DRIVEN NON-EXPERT DECISION MAKING

How do different types of data presentation impact non-experts' ability to make expert-like decisions?

Master Thesis
Copenhagen Business School 2019
MSc Business Administration and E-Business

Students:	Ulyana Khoma 121719	Csaba Vendel Poti 121699
Supervisor:	Rob Gleasure	

Page Count:	94
Character Count:	171 800

Abstract

As the rapid technological advancement shape our world, the importance of data and its transformation into information and knowledge through Business Intelligence or various other data analysis tools has become a widely researched area (KMPG, 2018; Wu, 2000; Niu et. al., 2009). However, only a few researchers chose the perspective of decision making, more specifically, the role of humans in the data-driven decision-making process. According Kahneman's (2003) theory of the dual process thinking system, human beings do not behave rationally. Using System 2 requires cognitive effort, the volume of which is determined by the individual's domain specific knowledge according to cognitive load theory by John Sweller (1988). Overall less experience in a domain could lead to higher cognitive load, which would then result in a lower decision quality. The focus of this research is to investigate this process and to further understand how to help novices understand data and make decisions based on it like experts would. The aim is to find specific practices in the presentation of data, which can be used by companies to help data novices, to understand data and enhance their learning processes by reducing their cognitive load in the Business Intelligence setting.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 General introduction

Data has become one of the most important currencies of the modern world (KPMG, 2018). However, a large volume of data in itself doesn’t translate into information or knowledge without the interaction of human cognitive functions (Wu, 2000). Decision support systems (DSS) are information systems, which are designed in order to minimize the cognitive errors and maximize the performance of decisions by the responsible persons (Niu et. al., 2009). The components of data science, data storage and analysis have evolved to a sufficient level for the focus to shift towards cognitive functions to support decision makers. (Medley, 2016).

Kahneman in 2003 stated that humans do not behave rationally, by describing decision making as a dual process system. Kahneman was the first person to place this view into an economic perspective. The dual process theory states that decision making can be processed in two different levels: One is the more intuitive and emotional (system 1) and one that analyses the decisions in a rational way (system 2). This process of system 2 requires mental capacity, which creates a cognitive load in the decision maker. Therefore, we have an automatic intuitive mechanism in our brain that tries to minimize the effort but make the best decision with the available resources (Kahneman, 2003).

High amount of cognitive load can disturb the use of system 2. It can be connected to learning processes. Cognitive load is generated by several components of novel information, such as the complexity of the dataset, the amount of redundancy included in the dataset, and the actual learning process itself. It makes individuals less alert and more prone to mistakes (Kahneman, 2003; Stanovich and West, 2000). Cognitive load theory by John Sweller (1988) is built on this idea, it states that domain specific knowledge is the primary factor distinguishing experts from novices in problem solving skills. Therefor in their area of expertise, experts generate a lower amount of cognitive load, which enables a better use for system 2, which results in a better decision in theory.

1.2 Scope

This research considers Business Intelligence as a starting point. BI involves decision making on all levels of a company ranging from operational short-term decisions to long term company strategies and including structured, semi-structured and unstructured decisions. (Niu et. al., 2009). As technology is evolving, data analysis becomes available for a wider audience with lesser experience in the area of data science. This would mean, according to the theories, a lower average in problem-solving skill, or decision quality. The research builds on this comparison of novices and experts, which states that experts on average have lower cognitive load and thus better decision quality. The aim is to further study this relation and unveil secondary factors in this equation. Can domain specific knowledge be substituted by any other factors in order to keep the decision quality? Can any characteristic of data be improved in a way to compensate or support novices to make better decisions or is experience an inevitable requirement for good decisions?

1.3 Motivation and contribution to society, business and academic audience

Many previous researches have focused on the topic and measurement of cognitive load, but limited number of previous papers have worked in the context of Business Intelligence and decision quality before. Examining the main statement of Cognitive Load Theory and constituting other factors can also be considered relevant for many areas and in time.

Amongst student, who are on the verge of finishing their studies and entering the job market, it was often a sore spot, that many positions require several years of experience. This paper glimpses into the relevancy of that experience and aims to provide alternative solution for companies to compensate for the lack of that experience. The contribution to academic audience can be derived from the method, as another research has tried to measure cognitive load using a special device to further help to understand the human cognitive architecture. The focus is to help companies to design their databases in a better way that benefits all participants of the process.

1.4 Personal motivation

The research has provided several new insights on a personal level as well. Looking at a mostly traditional aspect of business in business intelligence from a different perspective broadened our horizons. Using quantitative method and measuring physiological reaction through medical and scientific equipment let us glimpse into the life of professional researcher as an alternative to the more traditional job market.

Through working with experts, we had the opportunity to look into how professional experience can change us in many different ways and provided a way to look at experience as a job requirement in a different perspective, by providing theoretical background to it.

1.5 Research question

The researched topic has an immense depth and width and is worth the time spent on it for research. The scope was further narrowed down through a research question. The following question guides the available theories and methods used for the research in order to answer it in the conclusion, which is the following:

How do different types of data presentation impact non-experts' ability to make expert-like decisions?

1.6 Outline of the thesis

The thesis starts with this short introduction to provide a general overview of the research area and the related theories shortly. The next chapter, literature review provides the theoretical backbone of the paper, introducing Kahneman's two systems, John Sweller's cognitive load theory and a general introduction to human decision making and business intelligence, and what is the connection between the theories and decision making in this setting. The methodology starts by introducing the philosophy and the research method through Sander's (2007) research onion. Next the context of the research is discussed, where the dataset is coming from and what is the exact process of the experiment and as a final part the data analysis method is described. The Findings are described in three parts, descriptive analysis, where the measurement process and data outputs are discussed, then in a bird's eye view, the findings are represented on graphs. The second chapter contains the hypotheses, where the statistical data analysis is described by each hypothesis. The third part is the post hoc analysis, where further calculations and different perspectives of the data were examined. The discussion starts with a brief overview of the findings and the hypothesis are each discussed in light of the theory and the primary data collectively. Discussion ends with the limitation of the research and implication for further research. The paper is then briefly concluded in the conclusion containing the main findings.

2. Literature review

This chapter provides an overview of the existing theories, literature and academic research with regards to decision-making, business intelligence and cognitive load. Since this study focuses on cognitive load in BI decision making, it is first important to understand the foundations of what a decision is, how the decision-making process works, how humans make decisions and ultimately what factors influence the quality of a decision. It then moves to a discussion of how information and data evolved, followed by an explanation of BI and BI processes. Followed by this, a section is explaining the cognitive load theory and the fundamental principles related to it. Finally, a conceptual model derived from the literature review is presented along with the hypotheses that have been developed for this study.

2.1 Decisions and the Decision-Making Process

2.1.1 Decisions

The short definition of a decision, as described by Tarawneh (2012), is that it is a solution to a particular problem that has been chosen and will be implemented. In the long version Tarawneh (2012) defines a decision as follows: “an ongoing process of evaluating alternatives related to a goal, at which the expectation of decision maker with regard to a particular course of action impels him to make a selection”. In fact, he further specifies that it is a conscious choice to act or think in a certain way at a certain moment.

Furthermore, Tarawneh (2012) distinguishes between two types of decisions: formal and informal. The former is characterised as being uncommon and complex decisions; “policies, procedures, criteria, and methods for making such decisions may not always exist since the problem faced may lack precedent”. The latter is a more familiar and ordinary type of decisions; that are easier to make due to the methods have been put in place to aid decision makers. The following chapter looks further into the decision-making processes.

2.1.2 The decision-making process

Essentially, decision-making is the process of making a decision. According to Juslin & Montgomery (2007) it is a "constructive activity that aims to prepare the individual for effective action". More precisely decision-making is defined by Tarawneh (2012) as "the process of identifying and selecting from among possible solutions to a problem according to the demands of the situation". Bhushan & Rai (2004) argue that in today's ever changing and dynamic world decision-making is a great challenge. In fact, a decision-making process often involves uncertainty and ambiguity (Schwartz et. Al., 2011). Schwartz et. al. (2011) point out that people are prone to make mistakes when evaluating available information, making decisions and evaluating the outcomes of those decisions. Due to this, tools and processes have been created to help people make decisions. For example, University of Massachusetts Dartmouth (2019) provides a seven-step process to ensure effective decision making. The seven steps are as follows:

1. Identify the decision
2. Gather relevant information
3. Identify the alternatives
4. Weigh the evidence
5. Choose among alternatives
6. Take action
7. Review your decision and its consequences

It is explained that one should start with clearly defining the problem and the decision that needs to be taken. Then the necessary and relevant information needs to be collected from good sources. From there several potential courses of action will emerge; one of these will be selected after a careful evaluation of the gathered information and the person's feelings. Finally, the selected option will be implemented, and its outcomes will be evaluated to aid in future decision making. (University of M., 2019)

Furthermore, it is important to distinguish between risky and riskless decisions (Kahneman & Tversky, 1983). They characterise the level of risk of a decision by looking at its potential outcomes and the probabilities of those outcomes. For example, a risky decision could one where the decision maker accepts a gamble of receiving a monetary outcome under certain probabilities (Kahneman & Tversky, 1983). They argue that an individual makes such a decision without the preliminary knowledge of its consequences, since this type of decision usually depends on unpredictable events/factors. At the same time a riskless decision is described as a transaction in which a service or a product is exchanged for money (Kahneman & Tversky, 1983). Moreover, they state that the same decision could be perceived by a person as risky or not risky depending on how it is framed; whether it is framed in terms of losses or gains. Finally, Kahneman & Tversky (1983) suggest an additional factor that causes uncertainty in decision-making, and that is the mismatch between decision values and experience values. Kahneman & Tversky (1983) differentiate between people who prefer to make risky and risk-free choices. They characterise those who prefer a definite outcome over an uncertain one as risk averse. At the same time, those who favour uncertainty as risk seeking.

2.1.3 The Human Behind the Decisions

Assuming that human beings are rational beings, as is commonly believed, they should ideally follow a similar process as described above and ultimately come up with a perfect solution. However, Juslin & Montgomery (2007) argue that rather than searching for the optimal solution, people generally tend to "stop analysing at some more or less implicit outcome of a decision, when they feel they have achieved something useful from the problem-solving process". Moreover, in an experiment they conducted Juslin & Montgomery (2007) discovered that when given a cognitively demanding task people will tend to "find a way of performing a task that they cannot perform in an optimal way in a reasonable way instead...[and] finding information that simplifies the task for them.'' In other words, people will tend to "avoid going into complex forms of thinking when simpler information was available".

Johnson (2014) argues that the human perception is biased, and is influenced by past experience, by the context of the present situation as well as the present or future goals. In fact, he points out that people's attention tends to be affected by their goals, meaning that this for example will cause them to direct their attention to those goals. This happens because the "brain can prime [the] perception to be especially sensitive to features of what [the person is] looking for" (Johnson, 2014). Moreover, he argues that this causes people to inadvertently ignore other stimuli around them. By definition, a decision-making process requires that a person analyse relevant to their goal information. However, as Saaty (1990) argues, an individual that is not familiar with the subject is likely to have trouble deciding which factors to take into consideration and which not to. This could mean that, if one does not know what they are looking for exactly, they may have trouble paying attention to the correct stimuli.

In his book, Thinking, Fast and Slow, Kahneman (2013) argues that people generally make predictable, systematic errors when it comes to judgement and choices. In order to understand why such errors occur, it is important to understand how people think and how the human brain works. The book focuses on two ways of thinking inherent to people System 1 and System 2 as Kahneman (2013) refers to them. The former, is a fast, automatic system that takes minimal effort and is often referred to as the intuition. The latter, on the other hand, is a slower, more conscious system, that requires effort to engage in more mentally demanding activities, such as computations and making difficult choices. While it is common to believe that people mostly use System 2, the opposite is true. In fact, Kahneman (2013) argues that "the division of labour between System 1 and System 2 is highly efficient: it minimises effort and optimises performance". System 1, however, is prone to make systematic errors; it is always active and fast to react to stimuli which makes it susceptible to cognitive illusions.

Building up on Kahneman's theory of fast and slow thinking it can be argued that experienced managers make intuitively better-quality decisions. Having worked for years in a certain field increases the likelihood that the individual came across complex, formal decisions, where he or she had to engage System 2 in order to produce the best solution. This creates knowledge that is then stored in long term memory, and is automatically and almost intuitively retrieved by System 1 when necessary.

2.1.4 Strategic Decision-Making

Decision making in a company occurs on a regular basis, on every managerial level and involves both individual decision making, and group decision making. Strategic level decision making can be characterised by having very high stakes, "human perceptions and judgments are involved and whose solutions have long term repercussions" (Bhushan, 2004). Niu et. al (2009) distinguish between three classifications of decision problems: structured, unstructured and semi-structured decision problems.

- Structured decision problems are described those that can be explained using standard mathematical methods, such as statistics or linear programming. Solutions for them are relatively easy to find using traditional commonly known methods. This can be compared to the informal, routine type of decisions mentioned above.
- Unstructured decisions are described as "fuzzy, uncertain and vague, to which there is no standard solution method". This type of decisions could be compared to the formal, unusual type of decisions.
- Semi-structured decision problems are in the middle between structured and unstructured. "Solving this kind of decision problems involves a combination of both standard optimized solution procedures and human intuition or judgments" (Niu et.al, 2009).

Bhushan (2004) further mentions that there are several important questions that need to be considered when making decisions in a company, such as: "what decisions must be made, who will make them, how and what resources will be allocated, and how the situation will be measured and revisited in the dynamic environment in which the system will be operating". Moreover, he highlights that it is vital that guidelines and governance structures be put in place with regard to decision-making in large companies.

Numerous tools and processes have been devised to support and improve decision-making at the workplace. One of such processes is called the Analytical Hierarchy Process. As Bhushan (2004) defines it, "it is a systematic approach developed in the late 1970s to structure the experience, intuition, and heuristics-based decision making into a well-defined methodology on the basis of sound mathematical principles". It is a quantitative approach to decision making, in particular useful for unstructured strategic decision problems (Bhushan, 2004). In this process the goals, attributes, issues, and stakeholders are arranged in a "hierarchical structure descending from an overall goal to criteria, sub criteria and alternatives in successive levels" (Saaty, 1990). Additionally, Saaty (1990) highlights the advantages of this process; he states that "it provides an overall view of the complex relationships inherent in the situation; and helps the decision maker assess whether the issues in each level are of the same order of magnitude, so he can compare such homogeneous elements accurately". An important consideration to make, is that "to a person unfamiliar with the subject there may be some concern about what to include and where to include it" (Saaty, 1990). This means that the people using this approach may require assistance.

2.2 Decision Quality

According to the definition of decision-making form the business dictionary (BusinessDictionary, 2019) "for effective decision making, a person must be able to forecast the outcome of each option and based on all these items, determine which option is the best for that particular situation". However, as previously mentioned, people tend to make mistakes when evaluating available information and making decisions (Schwartz et al, 2011). Therefore, it is important and necessary to look at and discuss the various criteria that could be used for determining the quality of a decision.

Substantial research exists with respect to what factors could influence decision quality as well as the potential outcome. Some of these factors, that are most relevant for this study, include information quality, information load and overload, and even certain qualities of the decision maker.

2.2.1 Information Quality

With regard to information quality Raghunathan (1999) argues that it is common to believe that as information technology (IT) increases the quality of information, the decision quality increases and in turn the company performance.

In order to measure information or data quality Pipino et. al. (2002) suggest that it is necessary to keep in consideration both the "subjective perceptions of the individuals involved with the data, and the objective measurements based on the data set in question". Subjective perceptions mean that the decision maker's behaviour will be shaped based on whether he/she believes the information/data is of good quality (Pipino et. al., 2002). They argue that this is, however, something which can be measured through a questionnaire or a survey. In terms of objective measurements Pipino et. al. (2002) distinguish between task-independent and task-dependent assessments. The former refers to metrics which "reflect states of the data without the contextual knowledge of the application, and can be applied to any data set, regardless of the tasks at hand". The latter take into account "the organization's business rules, company and government regulations, and constraints provided by the database administrator".

When it comes to numerical data, Janssen et. al. (2016) discuss the effect of Big Data quality on decision quality. They argue that Big Data "is collected from different sources that have various data qualities and are processed by various organizational entities resulting in the creation of a big data chain". This could consequently result in noise and heterogeneity of the data and ultimately poor quality (Janssen et. al., 2016). They give an overview of the various existing methods and processes of Big Data collection, consequently highlighting the importance of knowing how the data was collected and who was involved in the data collection process. Furthermore, they found that the larger and the more complicated the data-set becomes the more difficult it is for people to interpret and comprehend it due to the "limited mental capacities of humans" (Janssen et. al., 2016). Additionally, information quality could also refer to the visual quality of the data presented to the individual. Janssen et. al. (2016) point out that the decision maker "should not be manipulated by fancy graphics" since "the ability of decision-makers to understand the data...results in better decision quality".

2.2.2 Information Load

Given that the information is of good quality, it could be argued that the more information a data set holds the better the decision will be, since the decision maker will be more informed. However, as research shows it could actually have a negative effect and deteriorate the quality of the decision. The reason for this is the concepts of information load and information overload. Jacoby (1977) defines information load as the amount and range of stimuli an individual can pay attention to. Moreover, Evaristo et. al. (1995) define information load as "a continuum in which both extremes are detrimental to human performance". He argues that presenting an individual with too few stimuli creates information underload, while presenting them with too many creates information overload.

Sweller (2002) argues that it is highly important to know the human cognitive architecture, in particular how people process information visually. When people process information, they use their working memory which is limited in terms of its capacity and duration (Sweller, 2002). These limitations however, "apply only to novel information that needs to be processed in a novel way" meaning that when new complex information is being processes long-term memory and learning mechanisms become engaged (Sweller, 2002). Should this effect take place in a decision-making situation it could have a negative influence on the quality of the decision.

Furthermore, Sweller (2002) presents several effects that different types of visualisation can have on people's cognitive load. For example, "the split-attention effect" occurs when certain information is incomprehensible without a visual presentation; this effect uses extra resources from the working memory thus overloading it. Sweller (2002) suggests that "the working memory load can be reduced by physically integrating diagrams and statements". Another effect he discusses is "the redundancy effect" where several types of redundancy exists. Two of the types of redundancy Sweller (2002) presents include visual redundancy (repetitive texts, descriptions and graphs) and redundant activities (need to engage in additional mental or physical activities, such as using a computer). The redundancy effect can increase cognitive load and as a consequence decrease decision quality.

2.3 Business Intelligence and Data-Driven Decision Making

This next chapter is going to present the theoretical background of Business Intelligence (BI) in order to help understanding the research question. First the evolution of data and information separately then evolution of BI is shortly introduced, then it is defined through various scholarly sources to come up with the description closest to the topic from a theoretical perspective. The following subparts describe the different elements, levels, specifics of BI and the competitive advantages and other benefits and values. To cater more to the question of the paper, the presentation methods of data is described generally in BI tools and in the final part of this description is the cognitive side of BI.

2.3.1 Information and data evolution

The creating, keeping, communicating, and consuming of information has always shaped the world in many different ways. Collecting data has been happening since the ancient times, but arguably, up there with writing, digital technologies have brought the biggest change for this process. (Gutcheckit, 2018) Invention of computers created a clear need for information and data processing. Originally, computer scientists had to write custom programs for processing data. The next step in the evolution brought assembly as a programming language, and Fortran, C and Java closely followed. The next natural step in the evolution was the invention of databases and SQL (Structured query language). These improvements broadened the possible audience of data analysis, as SQL language is read similarly to English and programmers weren't always required anymore for lower complexity tasks. (Krettek 2019) But the evolution still hasn't reached its peak. The next history defining change was the widespread use of internet. Exponentially more data, data types and new ways of collecting, using and analysing emerged with the introduction of internet. (Press, 2015)

All these steps have increased both the need and the possibilities of changing from analogue to digitized data storage. This process is called digitization. It is not to be confused with the definition of digitalization, which is the next step of the digital transformation process. Digitalization, as a major trend (Gartner Inc, a), 2017), is defined as "the use of digital technologies to improve the existing business model and create new revenue and value adding opportunities” (Gartner Inc, b), 2017).

The improvement has not stopped at this point. These technological upgrades brought opportunities for the ever-growing phenomena, Big Data. The huge volume of data enabled the industry of data science. Now that the data is available, cognitive technologies are surfacing in the forms of products that leverage AI to perform more complex tasks through machine learning and natural language processing. With technology advancing and trends are changing to cater to cognitive studies, human cognitive ability is replaced surpassed by technological advancement (Press 2015).

2.3.2 Evolution of Business Intelligence

The first source that specifically mentions Business Intelligence can be traced back to 1958 titled „A Business Intelligence System” written by Hans Peter Luhn (1958). The transition from paper to digital created the need for digital warehousing, which spawned the creation of the first iterations of database management systems. (Heinze, 2014). Technology has evolved rapidly ever since, data storage has increased in terms of volume and capacity, reaching the point of today, when storing terabytes and exabytes of high complexity data, such as different sensor or social media data, is a simple everyday task. This huge amount of high complexity data is referred to as Big Data (Conrad, 2019). In the early days, data analysis was a complex process, requiring experts in the area, but the newer vendors started catering to the less experienced or novice users. The evolution of processing speed enabled companies to make a decision based on real-time information through the connectivity of the business. BI became a requirement instead of a little addition for companies. (Heinze, 2014). Data has become the new currency in the business world, and analytics comes hand in hand with it. (KPMG, 2018). Hal Varian (2008) commented on the importance of evolving data analysis the following:

„So, what’s getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So, my recommendation is to take lots of courses about how to manipulate and analyse data: databases, machine learning, econometrics, statistics, visualization, and so on.” (Hal Varian, 2008)

The technology of our time has enabled us to further improve the capabilities of BI. With the introduction of Cloud and Mobile technologies, the availability and process speed is ever increasing and BI is reaching a wider audience.

2.3.3 What is BI?

Business Intelligence can be defined from many different perspectives. There is no single universal definition for BI: According to various authors and professionals and their definitions, the interpretation of BI consists of two important concepts in Data and Value. On a theoretical level Laursen and Thorlund (2010) simply defines the goal of BI to „deliver the right information and the right knowledge to the right people at the right time". Yeoh (et. al., 2008) emphasizes that BI improves both the quality of the decision and also improves the timing. From the technical perspective BI is a data-driven Decision Support System (DSS) designed to control the huge volume of data stored in data warehouses. BI is capable of extracting, transforming, loading and analysing the business data to support all levels, from operational to strategic, of the organization. (Conrad, 2019)

Data Warehouses (DW) are also worth a mention in this section. Kimball (2007) states that Data Warehouse is the foundation of BI. DW provides the data storage function and it enables BI applications to function properly and to gain an understanding of the data. A low maturity level DW is built up from Data Marts, which are independent single solutions in specific units of the business (Kimball, 2007). The other end of the scale is the Enterprise Data Warehouse (EDW), which is an interconnected centralized system of all the BI functions in a company. EDW enables higher level of data analysis which further encourages decision makers to rely on the information and make data-driven decisions (Rosedahl 2016).

2.3.4 BI decision-making process

A typical BI decision process takes five steps, which are Data Sourcing, Data Analysis, Situation Awareness, Risk Assessment and Decision Support. (Niu et. al., 2009) The first step can be described as data collection from various business units, like marketing or finance, and data transformation and integration. In Data Analysis, the data is converted into information through various BI data analysis methods, such as reporting, modeling and visualization. Situational Awareness is the step of turning information into knowledge, it is the deeper understanding of the decision. The goal of the BI system is to increase situational awareness to provide decision makers a clearer picture of the decision situation. Today's businesses are operated in an extremely complex environment. Risk Assessment helps decision makers to foresee threats and opportunities and to react accordingly. The end goal of this whole process is to provide Decision Support the managers based on all the available data. (Niu et. al., 2009)

2.3.5 The values of Business Intelligence

After the short introduction of BI in general, it is appropriate to talk about the benefits of BI. What value does it bring to a company? It is really necessary for corporations to stay competitive? According to Rosedal (2016) BI has several outstanding benefits in the form of, improving internal coordination, predicting future trends, quickly providing valuable information to responsible staff members, which enables quick response to both internal and external problems. Huie (2016) states that BI can provide information about the working processes of a company, identifying potential new opportunities and problematic areas, which enables companies to react to them accordingly. Hillsberg (2017) defines the general values of BI in five points. First of all, it provides relevant insights, which enables a wild variety of opportunities ranging from the companies determining their market position to learning about relevant customer behaviour. The second point is, it provides a bird eye’s view. It can give an overall picture of the company or just a specific area of it. Dashboards and scorecards are the relevant tools of BI for this area. Third, it centralizes data, which enables well-oiled organisations to save time by having one central source of information. Fourth value is the streamlining of processes. By automatizing the analytical processes, the focus can shift from calculations to decision making processes, providing the best opportunity for the responsible person. Various tools of BI applications are focused on this task, statistics, predictive analysis, computer modelling, benchmarking and several others. The fifth and last benefit according to Hillsberg is easy analytics. Modern BI softwares enable employees or even non-professional to make analytics easily by themselves without having years of previous knowledge in that area.

Decision are made throughout all levels of a company ranging from operational decision < managerial decision < strategical decision, such as decision from a product manager < site manager < CEO (Anthony, 1965). BI operates on every level and different decision levels have different techniques, tools and different requirements. As mentioned above, decisions are made on various levels, but the literature mostly focuses on the strategic and operational levels (Rosedal, 2016; Williams and Williams, 2007) Rosedal states *„BI is unique in that it has the potential to generate both 1) strategic and 2) operational value through the seamless integration of organizational data to support decisions at different levels of the business."* Rosedal (2016)

2.3.6 Maturity levels

In order to measure the data utilization or maturity of companies Wu (2000) came up with a model, where a higher maturity means better insight and more value for the company. He states that properly utilizing data, meaning high maturity level, can be a competitive advantage. The levels of maturity according to Wu are the following: Data, Information, Analytic, Knowledge, Wisdom.

Data level of maturity is described, as the lowest level, the collection, processing and storing of data. For example, if a supermarket sells a product they offer, they will store the item itself, quantity, price, cashier who sold and so on. The increase in the volume and the depth of the stored data increases the quality of it. This accumulation of data into a meaningful context provides information, which is the next step in the maturity levels. While this combination and meaning creation is useful, it just scratches the surface of the potential of the data. Transforming and regrouping information can extend the value of the information. OLAP (Online analytical processing) systems provide the user the opportunity to analyse information and determine relationships, patterns, trends and exceptions in the stored data, which is considered the analytic level in the model. Knowledge is the next level of understanding. It is defined by the interconnectivity of the previous levels, unveiling deeper connections between the layers of data, using various data mining techniques which are based on statistics and algorithms to provide users with the ability to discover knowledge within their data. The peak of this model is Wisdom. It requires the highest level of understanding in terms of data connectivity. The goal of this level is to use technology to reach, challenge and surpass the limits of the human cognitive abilities. Artificial intelligence, or machine learning to be more precise, aims to outperform the human capabilities of understanding and using data. (Wu, 2000)

2.3.7 The analytic tools of BI

The analytical tools are probably the most important fundamental core of every BI system Evelson et. al. (2008) summarized BI analytics into 8 different categories based on the popular BI tool options on the market.

Operational reporting for mass report distribution functions as the heart of BI. The goal of this tool is to distribute the available information to all parts of the company. Ad hoc query tools are used to provide answers to occasional business questions, a quick solution for a quick problem, which usually just requires a short query. OLAP tools are used in order to provide a deeper understanding of the problem. It answers the questions „why” rather than „what”. Also known as „slicing and dicing” analysis. OLAP provides different aspect of the data based on different dimensions. Dashboards as an interactive visual user interface, which is designed to provide historical current and predictive information in a visual way, to ease the findings of trends and exceptions. BAM (Business Activity Monitoring) report on real-time data and process information. It is providing the necessary information to make real-time decisions based on the delivered business metrics. Predictive modeling tool to not only answer the questions „what” and „why” but also the question „what is happening next”. Based on various statistical models, this tool attempts to predict metrics in the future by using possible existing and future conditions. BI workspace is an environment where capable users can access all BI functionalities with the highest available freedom, without depending on IT. Guided BI search tool is designed to enable less experienced people to get the same information from a dataset through search keywords.

2.3.8 The presentation of data in BI tools

As previous chapters suggested, data is the central part of BI, but the way this data is presented also plays a crucial part in value creation (Kumar, 2014). Data can be stored in various tools, as stated earlier, ranging from the top of the hierarchy, data warehouses or data marts, through different OLAP cubes with more dimensions to the more specialized specific reports, which all provide different solution to different BI related decisions. The types of reports have different layout to present the data, for example, Metrics, Charts or animated representation. Various applications are available for this purpose, some are more advanced and some are more basic to cater to users of all levels of experience. Tableau and Power BI are amongst the most popular tools (Scott, 2019), but even Microsoft Excel can function as one and is a popular choice for companies (Microsoft, 2019). The next step after choosing the reporting tool is the data display, which has many aspects such as the structure, the tone of the report based on the target audience, length, data structure, data positioning and even the data description and importantly, visual sophistication (Kumar, 2014).

*We are observing great change in data visualization; the initial visual data discovery releases from the big organizations like Microsoft, SAP, IBM, MicroStrategy, SAS, and Oracle tended to have limited capabilities, but the gap is slowly closing. (*Kumar, 2014*)*

With the industry evolving in such a steady rate, user expectations are also increasing. In the age of big data, companies have to face a huge volume of unstructured data to stay competitive, which requires sophisticated methods to analyse in a quick and accurate way. (Chen et. al., 2012)

2.3.9 Cognitive BI

As mentioned in the maturity model before, the goal of BI is to turn data into knowledge and then into wisdom in order to help managers to make decision, but currently most BI systems are working as data-driven decision support systems, they can only partially support managers' work (Singh et al. 2002). Ideally, a high complexity BI capacity combines the benefits of calculating power of a computer and the cognitive functions of a humans, but currently the latter is lacking, and just more data in itself doesn't provide more valuable information (Endsley et al. 2003) The pre-defined reports provide an efficient and quick solution to well-structured problems, but for ill-structured events, which would provide deeper understanding to more complex problems, they are not flexible enough. The trend of conscious design is emerging, as previously many considered BI as a time sink (Grumman, 2007). Resnick (2003) criticized the focus on executive dashboard design's improvement of data analytics, instead of the cognitive engineering considerations. As BI is getting more widespread, novice user's accessibility is increasing, as more and more BI solution provider focuses on expanding the BI user base. (Smuts et. al., 2015)

2.4 Cognitive load theory

The next chapter introduces Cognitive Load Theory (CLT) as the guiding theory behind this current research and explains why it is relevant for the question at hand. First, by briefly introducing the development of the theory, then emphasizing on its growing effects on problem solving and learning processes. As surrounding theories are very much interconnected, Cognitive load is going to be clearly defined in contrast to mental load and stress, which terms are often used interchangeably in some literature. Next, the content of the theory is going to be discussed, which are the components of the theory, what do they mean and what purpose do they serve. In the end Cognitive Load Theory is connected to Business Intelligence, and more specifically, why is it relevant for this research paper.

2.4.1 The development of Cognitive Load Theory

The development of Cognitive Load Theory began with the publication of "Cognitive Load During Problem Solving: Effects on Learning" in 1988 by John Sweller (Moreno and Park, 2010). The theory began as an instructional theory based on our knowledge of human cognitive architecture (Sweller, 2010). In cognitive load theory, the main focus is on the physiological construct of cognitive load and the relation of an observable phenomenon of practical consequence, which is learning. It was based on the work of Moray et al. (1979), who studies a similar phenomenon in “mental load” as the difference between task demands and the person’s ability to master (Moreno and Park, 2010). Both of these topics were researched thoroughly on their own by many researchers, but connecting them was considered innovative at its time (Madden, 1962; Mead, 1970). CLT has mostly focused on how the objective characteristics of the task affect cognitive load and, in turn, learning (Moreno and Park, 2010). The only individual characteristic, in contrast of mental models, that is explicitly included in its theoretical framework is prior knowledge (Kalyuga, Chandler, & Sweller, 1998).

2.4.2 Learning process

The processes of human cognition constitute a natural information-processing system that mimics the system that gave rise to human cognitive architecture: evolution by natural selection (Sweller, 2010). Human cognition and biological evolution are similar in the sense that both concepts create novel information, store it indefinitely and actively or passively uses it over space and time. Cognitive Load Theory is a representation of the general human cognitive architecture and the instructional changes inspired by this stored information (Sweller, 2010).

From an evolutionary perspective, learning, or generally gathering knowledge can be gathered through two methods. This knowledge is either biologically primary or biologically secondary knowledge (Geary, 2007). Biologically primary knowledge is defined by evolving information through generations. These include the most basic social skills, recognising faces and basic methods for problem solving, listening and speaking in native language. This knowledge is generated unconsciously, effortlessly and with the simple motivation of being part of society. This whole process is also observed in the animal world. They learn how to engage in a social interaction with members of their own species and how to hunt and avoid predators, in most cases (Geary, 2007). For example, when a young child learns its native language, no one is teaching them specifically, on a fundamental level, how to use their lips and tongue to produce each specific sound used in its native language.

On the other hand, biologically secondary knowledge is culture dependant (Sweller, 2010). We have evolved the ability to gather such knowledge in specific ways, but not the specific knowledge itself. It is the opposite of primary, as in it takes effort and requires mental effort to gather. As society is not enough of a motivation for it, and its culturally different, institutions were created in order to enforce this kind of learning. Cognitive Load Theory is strictly relevant for biologically secondary knowledge, as only this type generates mental effort. As this knowledge is generated in specialized institutions, they are in the focus of the cognitive load theory.

Human cognition is governed by five basic principles (Geary, 2007) when dealing with knowledge and involves both biologically primary and secondary knowledge generation and as such constitutes as a natural information processing system and derive from evolutionary theory (Sweller, 2003, 2004).

The five basic principles are:

- Long-Term Memory and the Information Store Principle
- Schema Theory and the Borrowing and Reorganising Principle
- Problem Solving and the Randomness as Genesis Principle
- Novice Working Memory and the Narrow Limits of Change Principle
- Expert Working Memory and the Environment Organising and Linking Principle

The first principle in short version means, that human cognition has huge volume of information stored and it governs most of our activities, and long-term memory is basically the “hardware” where it is stored. The borrowing and reorganising principle states, that the memories stored in the long-term memory are borrowed from other long-term stores. For example, listening, writing or imitating another person is such a process. The randomness as genesis principle states, that for unfamiliar problems, where no previous knowledge and neither another long-term store is available a solution is randomly generated and then tested thereafter. This would mean, all knowledge is generated from this process originally. Narrow limits of change principle states, that changes to long-term memory happen slowly and incrementally. This is the case when new information is presented to an individual, for example school lasts for many years for everyone for a reason. The last principle is the environment organising and linking principle, which states that organised information from long-term memory can be used by the working memory without limit, to understand the connections of the complicated external world. (Geary 2007)

These five principles provide a base for human cognitive architecture. All five of the principles are strictly necessary for the system to properly work. The centrality of the five principles to human cognitive architecture mirrors their centrality to Cognitive Load Theory and instructional design. Each principle has instructional design consequences (Sweller, 2010).

2.4.3 Problem-solving

Sweller (1988) in his original Cognitive Load Theory paper starts by describing the importance of learning and complements with the statement, contrary to many cognitive theories, some forms of problem-solving interferes with learning. Problem-solving skills have always interested many scientists, especially in the area of mathematics and science (Sweller, 1988 and Dewey, 1910, 1916). Sweller differentiated people based on their previous expertise in the observed area, creating novice-expert distinction. This breakdown was sectioned into three categories: Memory of problem state configurations, Problem solving strategies and Features used in categorizing problems.

Memory of problem state configuration is originated from a study conducted by De Groot (1966), who compared chess masters and novices. The findings were different from expected. The superiority of the masters was not due to general short-term memory factors according to the test which was designed to prove this. This meant that the difference's source originates from long-term memory.

Problem-solving in general can be compared to problem-solving in mathematics. The problem is described with an initial state, a desired outcome and valid rule-following transformations between the two. There are different ways of getting to the solution from this point. Novices were found to often rely on means-ends analysis, which means that they try to reduce the number of unknown variables based on the expected outcome, and in the end use the original processing order to check the validity of their solution. Experts don't rely on the backward-working phase and solve problems by recognising appropriate equations leading to the goal.

These cognitive structures are called schemas, where a particular problem can be categorized by the particular move required to solve it. Novices, not familiar with these schemes are stuck solving problems through means-ends analysis.

Features used in categorizing problems is based on the previous statement. As stated, problems can be categorized by their solutions by the experts, as they are familiar with the schemes required to do so. In contrast, novices rely on the information they possess to categorize problems, which is the problem statement.

Overall, we can state that:

“Domain specific knowledge is the primary factor distinguishing experts from novices in problem solving skills.” (Sweller 1988)

2.4.4 Scheme acquisition

As stated in the last chapter, the possessed schemes are defining knowledge in specific areas. The next surfacing question must be related to scheme acquisition. Sweller (1988) states, that as means-ends analysis is providing the bad incentive (for example providing a math problem with the solution included, requiring only the solution method) and is also related to the novice side of the experience dimension can be a counterproductive effect on the efficiency of learning processes. He stated that, as it is also observed for experts, forward working problem-solving provides a more efficient way for scheme acquisition.

Sweller based his findings on the working memory model of memory by Baddeley and Hitch (1974). It explains the relation between the short-term (or working memory) and long-term memory. It can be compared to a computer’s RAM and hard drive in many ways. Working memory has a limited working capacity, which requires the learning processes to be as optimised as possible for maximum efficiency. Forward working is suitable for this as it only focuses on the variables and the next step, not the beginning and ending of the process, and what should be in between. (Sweller, 1988) Miller (1956) stated that the human working memory, when focusing on novel information, has a capacity of focusing on 7 elements at once. Sweller (1999) added that partially separate visual and auditory channels are not included in this number.

There are two methods to test the efficiency of the learning process. These methods are Recall and Transfer. The timing of this process can also vary. Immediate and delayed tests are available. Recall is suitable for testing facts, lists and processes. Conceptual learning is tested well through Transfer, where a previously explained scheme has to be applied to a different setting. (Sweller, 2006)

2.4.5 Components of cognitive load

Other factors can also be considered to optimize the capacity of the working memory. Sweller (1999) differentiated 3 categories of cognitive load. The intrinsic, extraneous and germane cognitive load, as presented in Figure 1.

Intrinsic cognitive load refers to the complexity of the novel information. The exerted load is higher if the task or concept is more complex. This cognitive load can be lowered by breaking down the process to several smaller and easier parts for the individual to complete. For example, element interactivity can create intrinsic load.

Extraneous cognitive load collects all the elements, which distracts the student from the original learning goal. This includes visual and auditory distractions and other redundancies. It can be reduced with effective case-appropriate presentation methods. For example, appropriate visual representation can help to explain various interactions.

Germane cognitive load is the process of scheme creating, the actual learning process. Both intrinsic and extraneous cognitive load should be kept as low as possible to encourage a higher proportion of germane cognitive load. This process imports the schemes from working memory to long-term memory.

The cognitive load capacity can be compared to a bucket. It has a limited capacity, and we want to fill it up, but this capacity is lower based on the amount of unfavourable dirt (Intrinsic and Extraneous cognitive load) it contains, and it leaves less place for the desired water (Germane cognitive load).

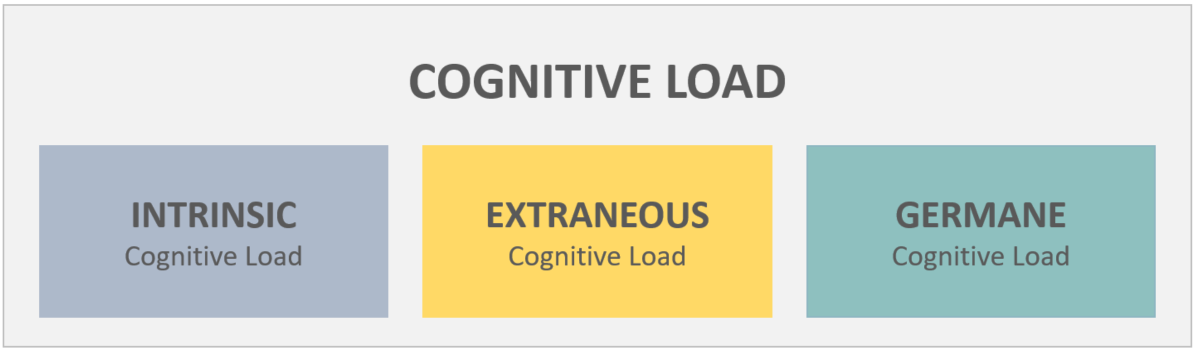


Figure 1: A diagram representing the three components of cognitive load

2.4.6 Differentiation of Cognitive load from other similar concepts

As Cognitive Load Theory is so interconnected with other researches in this similar area, it can often be seen that it is used interchangeably with other phrases. It originates from Moray's (1979) mental lode, which creates some confusion. DeStefano and LeFevre (2005) clears it up at the beginning of their research. Cognitive load is a construct with three measurable dimensions: mental load, mental effort and performance (Kirschner, 2002; Sweller et. al., 1990; Sweller et. al., 1998). Which puts mental load as only a component of cognitive load, what is imposed by task demands.

The other phenomena which is often mixed or used interchangeably with cognitive load is stress. Existing studies often use mental workload as stress-eliciting factor, however, there are other contributing factors such as social stress (Setz et. al., 2009). Which means cognitive load is a component of stress, which can be measured the same way as stress in a controlled environment. (Niculescu et. al., 2009)

2.4.7 Implications for the research

The most important statement of the theory is that domain specific knowledge is the primary factor distinguishing experts from novices in problem solving skills and as discussed in the business intelligence and data-driven decision making, technology is opening up for an increasing number of entrants to data analysis with lesser experience. This information in itself indicates that the average decision quality is decreasing, but as domain specific knowledge is indicated as the primary, not only determining factor, the question arises. What other factors influence problem-solving skills and in connection decision quality. Can they be used in order to enhance skills and hide the lack of experience of novices?

2.5 Linking the concepts of Data, Cognitive Load and Decision Making

The main concepts of this research were thoroughly researched in previous researches, but the added value is deriving from the combination of these concepts. The main concepts of this research are Data, Cognitive Load and Decision-making. The fundamental idea of the following model is that the observed Data has defining characteristics described below as the Amount of Data, Interpretability, Visual Representation, Completeness and Relevancy (Pipino et. al., 2002). These characteristics of the Data generate different quantity of Cognitive Load in the decision-maker, which later influences different aspects of the decision, namely the Accuracy, Feasibility, the Speed of the Decision and the Risk of the decision, (Boss, 2015; Parsons, 2016) which in the end impacts the quality of the overall decision. These characteristics of the data are defined in Table 1 on the next page.

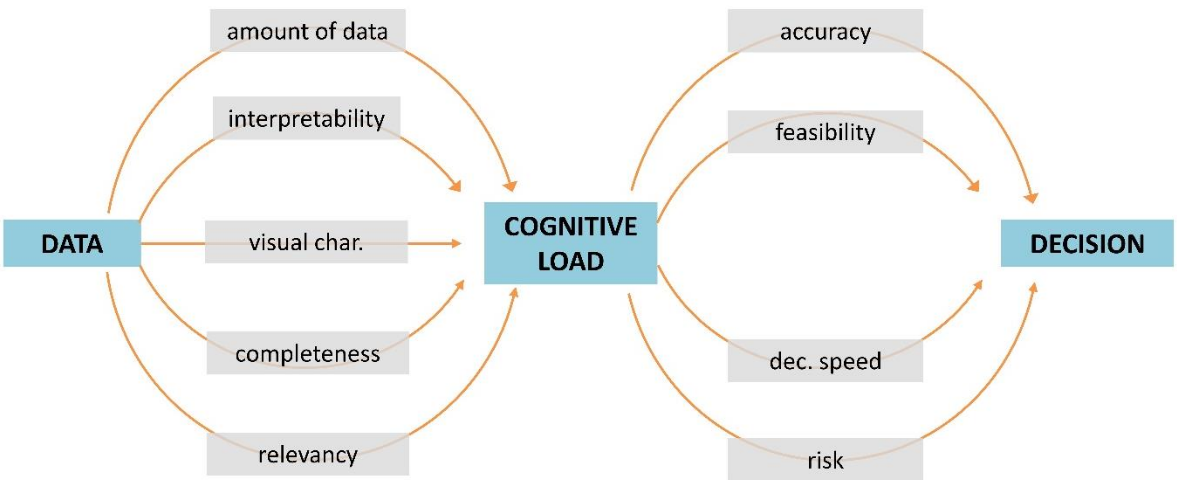


Figure 2: Overview of the connected theoretical concepts

CHARACTERISTICS	DEFINITIONS
Amount of Data	The volume of available and appropriate data for the relevant question.
Interpretability	The extent to which data itself and the symbols, units and definitions are easily understandable.
Visual Characteristics	The depth of visual elements and graphs available for the relevant data.
Completeness	The extent to which the expected data is available and sufficient for the relevant question.
Redundancy	The ratio of appropriate and redundant data.

Table 1: Characteristics of Data

CHARACTERISTICS	DEFINITIONS
Accuracy	Accuracy of the decision. Its net impact on different stakeholders.
Decision Speed	The time the decision-maker takes to make the decision.
Feasibility	The extent to which the decision is feasible in terms of time, funds, and scope.
Risk	Perceived/Assumed risk of the decision

Table 2: Characteristics of Decisions

2.5.1 Conceptual Model

A conceptual model can be defined as a set of relevant concepts, which represent a system. It helps people to know, understand or simulate a subject the model stands for and to understand the interaction between the different concepts that it consists. The term conceptual model is used to refer to models which are formed after a conceptualization or generalization of a process. (Tatomir et. al., 2018)

As a research of this magnitude is heavily limited in time and other resources and the overall scope of decision making deserves a lifetime to spend on researching, to understand the whole decision-making process with every determining aspect is not manageable and the most interesting perspectives had to be chosen, to be able to provide a more in-depth and interesting analysis on them. Out of the data characteristics, aspects closest to visualization were found interesting and were chosen as the focus of this study. A conceptual model was created in order to present this deeper connection. As the whole focus of the paper is to compare novices and experts, it didn't have to be specifically highlighted in the model. This model already aims to provide a basis for the hypotheses of the project.

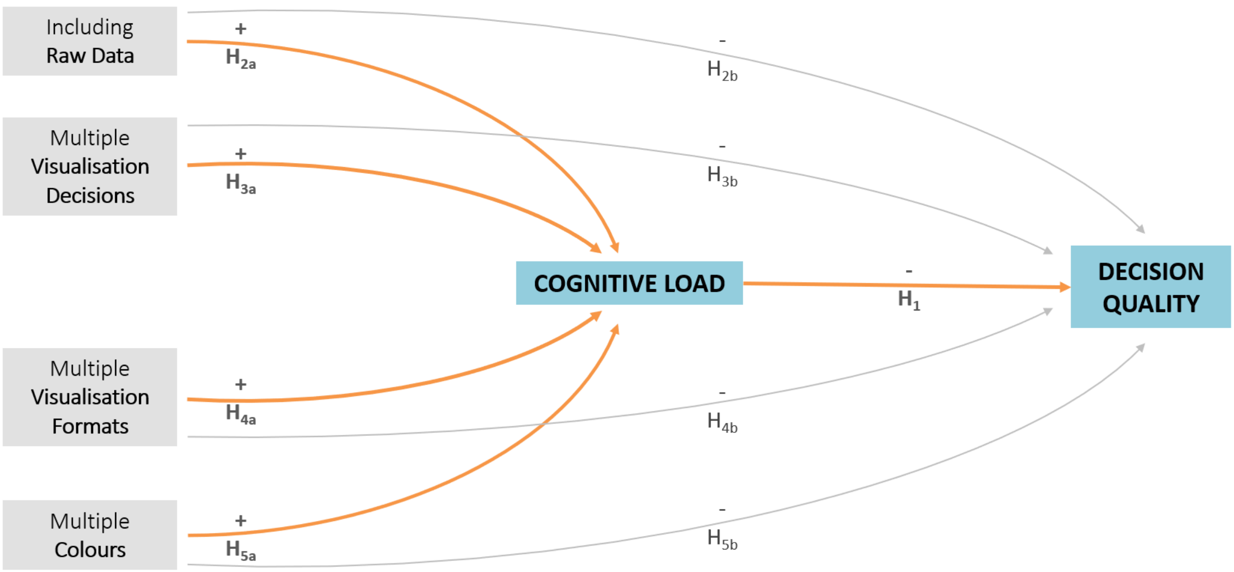


Figure 3: Conceptual model of the research

2.6 Hypotheses

The Conceptual Model in Figure 3 above displays the hypotheses that have been derived based on the reviewed literature. Hypothesis 1 focuses on the main statement of the research, which is that cognitive load influences decision quality. It is derived from the previously mentioned theories, Kahneman's Two System theory and Sweller's Cognitive Load Theory in an inductive way. It is necessary to examine the main statement in order to talk about the deeper analysis, which heavily builds on this statement. According to this, Hypothesis 1 was defined as the following statement.

> **H_1:** Increased cognitive load decreases decision quality

As the next step, attributes were identified, which could be examined in terms of cognitive load and decision making. Raw data, Visualisation Decisions, Visual Formats and Number of colours, as shown in Figure 3, are the four attributes that are predicted to influence cognitive load of the participants and both directly and indirectly the decision quality. This direct and indirect connection is represented in an "a" and "b" version for each hypothesis. The exact effect is described below for each of the hypotheses 2a to 5a respectively:

> **H_{2a}:** Including **raw data** increases cognitive load
> **H_{3a}:** Adding new **visualisation decisions** increases cognitive load
> **H_{4a}:** Adding data **visual formats** increases cognitive load
> **H_{5a}**: Increasing the **number of colours** in a data-set increases cognitive load

These statements refer to the direct impact on cognitive load. Additionally, another layer of hypotheses can be observed in the final conceptual model; namely hypotheses 2b to 5b. In these hypotheses the effect of the, previously mentioned, four attributes directly on decision quality are analysed.

H2 states that the inclusion of raw data in the document can increase cognitive load and indirectly cause lower decision quality. It is derived from both intrinsic and extraneous part of cognitive load. It can increase the complexity of the dataset, by providing unnecessary complex data with any different columns. On the other hand, it is often providing a better understanding of how the dataset is built up, so it is also dependant on the required depth of understanding for the task.

H3 expects, that bigger freedom in visualization decisions can increase cognitive load and indirectly lower the decision quality. It provides a way for decision makers to pivot the dataset to their own preferred format and their preferred perspective on the data. However, it can create the tendency for confirmation bias based on the chosen perspective, which can impact the decision quality.

H4 predicts that inconsistency of visualization formats can cause a higher cognitive load and indirectly lower decision quality. This statement can also be connected to extraneous cognitive load, which is impacted by the way data is presented. A consistent format provides an easier way to scheme connection ability, but can be skewed by personal experience of preference with the type of format.

H5 states that the addition of colours to the data itself can increase cognitive load and indirectly reduce decision quality. Colouring the dataset provides an extra layer of information to understand to the data, and is expected to provide a slightly better understanding, but the extra mental effort is expected to be utilized in a more efficient way.

3. Methodology

The methodology part contains the system of methods used for the particular research. The explanation starts with the introduction of the research philosophy, which determines most aspects of the methodology, arguing why is the chosen philosophy the correct one for the current research. Next Saunder's (2007) Research Onion is used to delve deeper into the main areas of the methodology, more specifically: the previously mentioned research philosophy; research approach; research strategy; research choices; time horizons and techniques and procedures. After the theoretical part, the practical follows. The setting of the research is presented to start the practicalities. Where the dataset originates from and how experts were managed to be connected to the thesis. The protocol of the experiment, the description of the survey and the case questions and the sampling of the participants for the test are all explained respectively, while data analysis is the final topic which is getting described in the boundaries of methodology.

3.1 Research Philosophy

According to Guba and Lincoln (1994) a research paradigm can be defined as a set of basic beliefs that deals with the ultimates or first principles, also known as research philosophy. It represents a worldview that defines, for its holder, the nature of the "world", the individual's place in it and the range of possible relationships to that world and its parts.

Guba und Lincoln (1994) stated that a research paradigm can be characterized through their ontology, epistemology and methodology. Ontology can be summarized with the question of: What is the form and nature of reality and, therefore, what is that can be known about it. Epistemology refers to the question of: What is the nature of relationship between the knower and or would-be knower and what can be known. At last, methodology is defined by the question: How can the inquirer (would-be knower) go about finding out whatever he or she believes can be known. More clearly, Patel (2015) explains that ontology means "what is reality?" epistemology "how do you know something?" and methodology "how do you go about finding it out?".

Based on these questions, different research paradigms are defined by different sources. For this research, the philosophies of both Saunders (2007) and Guba and Lincoln (1994) were examined to give the most accurate answer. The research philosophy, which represents the researcher’s opinion on the current paper the most was defined as, Critical Realism by Saunders (2007) and Post positivism by Guba and Lincoln (1994). The ontology of this paradigm states, that there is one reality, just like in regular positivism, but unlike in that case, reality can be known only imperfectly and probabilistically. In terms of epistemology, it states that human knowledge is based on human conjectures. As for methodology, as opposed to regular positivism, it builds on exclusion and instead of verifying theories, it focuses on falsification. It builds on various methods and different measurements and new ways to reflect on that one reality from different perspectives. This philosophy often focuses on mathematical data, but takes into consideration the context of individuals and experiments, which is crucial in order to understand social sciences.

3.2 Research Onion

Saunder's (2007) Research Onion collects the methodological approaches of research into six layers. The onion references the layers of the model, and the order of examination of the layers, starting from the outer, and working towards the innermost layer. The layers are called, following this order: research philosophy; research approach; research strategy; research choices; time horizons and techniques and procedures.

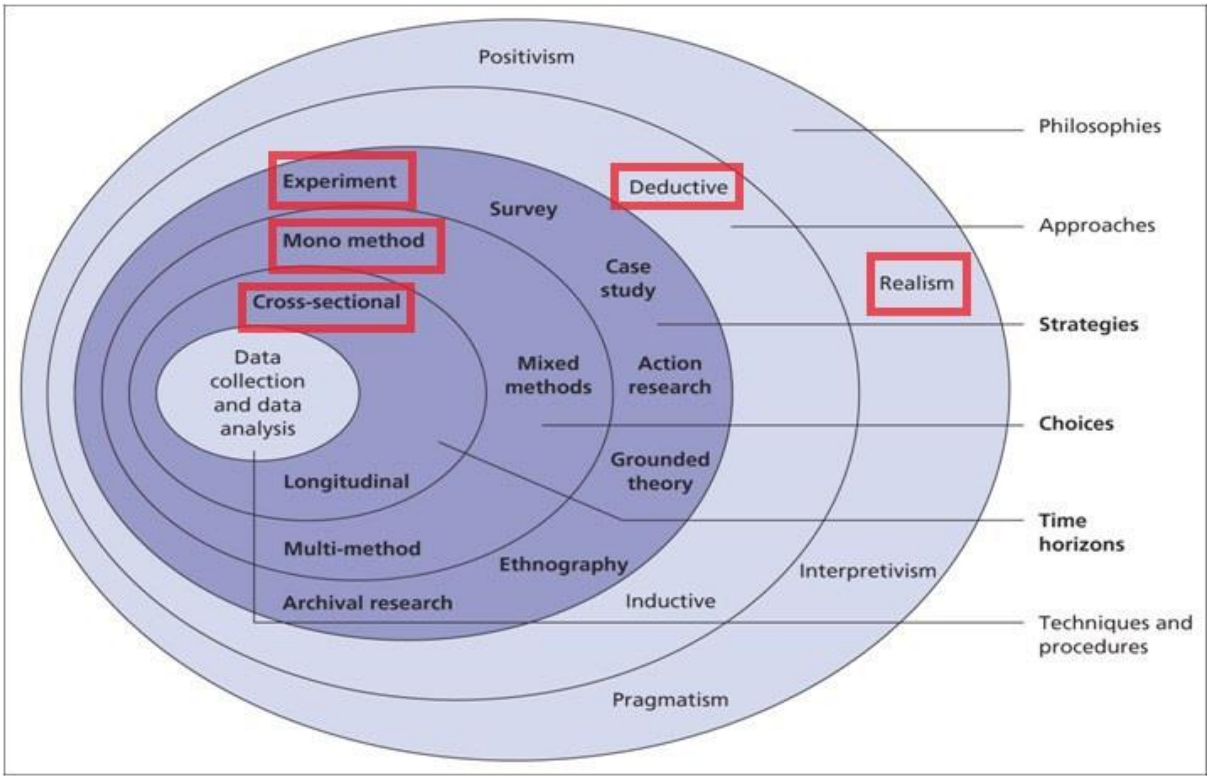


Figure 4: Saunders' research onion in this research setting.

The first layer, the research philosophy often determines several other layers, it deserves its own section in the paper. By Saunders it is referred to as Critical Realism and focuses on the one reality that is hard to grasp. In this research, the hard reality to grasp, is the optimal decision. Several aspects can be considered, but in this case, one good answer is the true good answer, and that is the expert's opinion.

The next layer is the research approach, which can be inductive, deductive or abductive. The chosen research approach for this research was deductive approach, or the top-down version of the previously mentioned. The reason to choose deductive is that, the paper builds on two important theories in Kahneman's two system and Sweller's Cognitive Load Theory. They provide the theoretical background, and through experiment, parts of these theories are tested. Although deductive is the main approach, some aspects of the research use different approach. Post positivism states the one true reality is hard to grasp, so it requires a throughout data analysis, which can present aspect in an abductive way, which were not expected.

The Methodical choice is a mono-method quantitative method. The designed experiment is albeit a complex process with many different outputs, can still be described as a mono-method, and both the survey part and the case questions, as well as the Cognitive Load readings produce quantitative data.

The research strategy, as mentioned in the last section, is an experiment. Several papers can be found on the topic of Cognitive Load measurement through galvanizer bracelets (Nourbakhsh et. al. 2012, Shi et. al. 2007). This process will be further elaborated in later sections of methodology.

The time horizon of the research is cross-sectional. Longitudinal would be both limited by the time, and as human learning capabilities are the outcome of thousands of years of evolution, the evolution of technology would be the only suitable as the time dependent variable, which would still require several years. Data collection and analysis will be further elaborated on in later chapters.

3.3 Research Setting

The research setting is to an extent defined by the fact that it is conducted as a Master Thesis for Copenhagen Business School. This determines the likely location of primary information gathering in Denmark, which has several implications on the physical, social and cultural setting of the research. One of the defining characteristics of Post positivism is that it doesn't require perfect laboratory context for experimenting, as the context also holds valuable information.

The topic was also determined by the context of school through supervision meetings, which enabled experimenting through devices provided by Copenhagen Business School, which can measure Cognitive Load.

Other implications of the context were also used in order to enhance the thesis. Through friends and family, the researchers could contact and work together with a company which prefers to remain anonymous. Initially their current active channel distribution dataset was shared with a research in order to design a credible experience for the participants. The file contained channel distribution in terms of traffic, conversion rates and transaction numbers. Later the experts of the company, who are working at the department, which works with this type of data, agreed to complete the finalized designed experiment, to provide insights on the test, and to help determine the good answers for the case questions, in order to determine the decision quality for the participants.

3.4 The Experiment Protocol

The following section describes the protocol of the experiment that was conducted for this study. Table 3 below summarises the four steps in which the experiment was conducted. This chapter follows the order presented in the table. It begins with describing the initial preparations before the experiment, then the description of the experiments itself. Finally, this is followed by a brief description of the hypotheses testing process as well as the additional tests that could be conducted to analyse the findings.

Step 1: Pre-Experiment Planning	• Data-Set preparation • Case Description Preparation • Survey Preparation • Equipment Selection & Testing • Participant Invitation • Location Selection & Booking
Step 2: The Experiment	• Location preparation • Equipment & software set-up • Execution of the experiments
Step 3: Data Analysis	• Data download & grouping • Basic Data Manipulations and calculations
Step 4: Post-Hoc Data Analysis	• Removing noise & outliers from data • Running statistical tests

Table 3: Structure of the Experiment Protocol. It displays the four steps in which the research was be conducted

3.4.1 Planning

The planning process began with preparing the case data-set based on the number from the original data-set provided by the company. Only selected parts of the original dataset were used, a formula was applied to all the values and finally sensitive words and titles were replaced as to ensure the anonymity of the company. Based on the studied literature the dataset was replicated into several presentation versions as to fit the hypotheses.

This process generated 16 Excel versions, with all the combination of hypotheses. The versions which contained the second hypothesis contained the raw data sheet in an unhidden version as the opening sheet. The third hypothesis deserved a slightly different survey, where more emphasis was placed on the availability of channel selection for the graph, and the necessary button for it was coloured in orange to make it more visible on sheet "Weekly". The fourth hypothesis refers to the consistency of data formats, these versions had three different visualization formats for the three graphs on sheet "Weekly". The fifth hypothesis referred to the colours, which were added in the version for all numerical data by Excel's conditional formatting option. Detailed differences are available in Appendix 1.

Then, based on the dataset partial instructions, case description and the case questions were clearly and concisely written down. The instructions were divided in two parts verbal and written; the former would be read out to participants before the experiment and the latter they would read in the survey. Finally, a survey was compiled in Google Forms, which was selected due to its ease of use and functionality. The case description and data-set are entirely fictional, but have both been inspired by a real-life e-commerce company that we refer to as Company X. All the values were altered as to ensure the anonymity of the company that wishes to remain unknown.

The case description was the following:

> *"Company X is a successful/large online supermarket. You have recently joined the Digital Marketing/E-commerce team of Company X as responsible for the traffic to website and conversion rates. Your boss saw the potential in you and assigned you to an important task. Looking at the data available to you, you will need to make several strategic decisions with regard to the digital development of the company.*
>
> *Each question comes in two parts multiple choice and an open answer, so we kindly ask you to fill out both parts. The excel file includes the following sheets:*
>
> *- Weekly (weekly traffic, conversions and transactions data for this year)*
> *- Long Term Data (weekly traffic data starting from January 2018)*
> *- 2019 Averages (average traffic, conversions and transactions for this year)*

- Responsibilities (distribution of working hours by channel and the responsible employee)

Feel free to look through all of them to help you make the decisions."

In the meantime, selecting and testing the equipment took place. For measuring the physiological response of the participants two Empatica E4 Wristbands were selected; the decision was made to use two devices to ensure certainty in the readings. Moreover, another decision was made to use the researchers' personal Android smartphones and Microsoft Windows Laptops due to convenience, ease of use and a limited budget. All the devices as well as the compiled survey and datasets were tested in a simulation experiment by the researchers and their friends to detect possible errors and malfunctions. The errors that occurred were immediately noted, corrected and the simulation experiments were re-run again before conducting the final experiments. Such iterations were done a limited number of times due to time constraints.

Finally, the participants were invited to take part in the experiments. Knowing the availability of the participants several conference rooms were booked at Copenhagen Business School campus. The rooms were chosen to be as similar as possible and in quiet locations to ensure that the conditions are as similar for every participant as possible.

3.4.2 The Experiment

As previously mentioned, the experiments were preliminarily scheduled and a suitable location for conducting them was found. The researchers always made sure to arrive well in advance to prepare the location. This included, for example, checking that the temperature in the room was not unusual and preparing a comfortable seating area for the participants, as to avoid increases in EDA for reasons other than stress. Additionally, the equipment needed to be checked and set up upon the arrival of the participants, leaving enough time to fix or charge whatever was necessary to ensure a smooth experiment performance. Moreover, in order to avoid any distractions several notes were written on an A4 sheet of paper and hung outside the experiment rooms, asking people around to stay quiet.

Both researchers were always present at all of the experiments; which were conducted in the following way. Upon the arrival the participants were greeted, and the researchers had a casual chat with them while they could remove their jackets and have a seat. They were then seated in front of the laptop and the following instructions were read to them:

> *"Hi and welcome. Thank you for agreeing to participate in our experiment, we appreciate it. We will now attach the wristband on your wrist and begin the measurement while we set-up the survey for you. This is the data-set you can use to answer the case questions, and this is the survey which begins with some standard survey questions. After these questions you will be directed to the page with the case description and further instructions. There will be five case questions that are divided in two parts: multiple choice and a window where you can explain your answer. The survey will conclude with a few additional standard survey questions You will have 30 minutes to do this exercise, but we will not interrupt you when you go above this time. You may begin now."*

The participants indeed would not be interrupted, as mentioned in the instructions, however when the time would reach 30 minutes the researchers would ask the participant how they are doing and how far along they are. Upon completion of the experiment the participants received a chocolate bar and were verbally asked how they felt and what they thought of the experiment.

3.5 Sampling

Sampling is an important process to go through regardless of the research question or the data collection strategy (Saunders, 2007). For the research question in this study, the decision was made to collect a sample rather than study the entire population. The target group of this study is Master level students who:

- will graduate in the next year or two
- are novice or amateur users of Business Intelligence
- are novices in terms of analysing and working with data in general.

Due to the size of the population of the target group analysing every possible case was not feasible. Moreover, the time constraint caused by the thesis deadline limited the sample size for this study to 32 participants. This not only allows for a "possibly higher overall accuracy that a census", but also reduces the amount of data that needs to be collected and consequently reduce the data collection and data analysis time (Saunders, 2007). On the other hand, it enables the research to provide a more in-depth analysis based on sample.

Selecting from various sampling techniques non-probability type of sampling was chosen. More specifically, a combination of purposive (judgemental) and self-selection volunteer sampling technique was chosen. The potential candidates were selected as the most suitable to answer the research question of this study, based on the researchers' own judgement. In the case of this research paper, the participants were expected to come to a specified location, therefore a decision was made to personally ask each of the selected potential candidates if they would be willing to take part in the experiment using a standard description of the experiment.

3.6 Measurement

This section looks at the data collection methods that were used in this research. An important distinction needs to be made between different types of data in this study. When talking about the data-set, the researchers refer to the final case data-set that the participants based their decisions/answers on. At the same time, when talking about data-collection, the researchers refer to the collection of EDA measurements of the participants as well as their answers. This distinction is to be applied throughout the entire paper.

In this experiment three dependent variables (dependent or mediating/moderating) are being measured, namely cognitive load, perceived cognitive load and decision quality. A combination of data collection methods was used to measure the three variables: questionnaires and observation. The former is defined by Saunders (2016) as a "general term to include all methods of data collection in which each person is asked to respond to the same set of questions in a predetermined order". While the latter, although sometimes neglected, involves "systematic viewing, recording, description, analysis and interpretation of people's behaviour" (Saunders, 2016).

Self-completed, internet-based questionnaire types were chosen to measure the participants' perceived cognitive load and their decision quality; it was chosen because the participants had to be physically present at the experiment location and could fill out their answers in a Google Form.

To measure the actual cognitive load structured observation and recording were used. Shi et. al. (2007) concluded that Galvanic Skin Response (GSR) "can be used to serve as an objective indicator of user cognitive load level in real time" and in fine detail. The researchers used two E4 Wristbands from Empatica to measure the galvanic skin response of the participants during each of their case solving processes.

Multiple recommendations can be found in literature as to how the GSR can be measured more accurately. For example, according to Braithwaite et. al. (2013) such recordings are "composed of samples acquired at discrete time points...described as the number of samples acquired per unit of time".

As mentioned earlier, cognitive load measurement is a complex process, as EDA has different components. According to Braithwaite (2013), EDA is built up from two main components of tonic EDA(SCL) and Phasic Skin Conductance Response (SCR). Tonic EDA is a constantly moving baseline, which suggests researches to conclude that Simply averaging across the whole signal is woefully inadequate as a measure of SCL because it likely to over-estimate the true-SCL as such averages will also contain SCRs (thus artificially elevating the measure) (Braithwaite et. al., 2013).

Moreover, they recommend setting this sample rate "quite low for long-term ambulatory measurements or experiments that do not require a high level of temporal precision (i.e., 1-5 samples per second)". Due to the time constraints the decision was made to use the standard sample rate of the E4 wristbands, that was approximately 1 to 3 samples per second. Additionally, they recommend "to run a baseline measurement period when the participant is not engaged in any given task" (Braithwaite et. al., 2013). Empatica, the wristband brand, themselves recommend the time to be of approximately five minutes (Empatica, 2019). While the participants had a few minutes to calm down upon their arrival, the bracelets were generally put on just one or two minutes before the beginning of the experiment. The baseline measurement period is, however, addressed in the data analysis section below.

3.7 Data Analysis Approach

The collected data was in its basic and raw form, which conveyed little meaning without processing and then analysing it (Saunders, 2016). Saunders (2016) states that in order to be useful the data needs to be turned into information, that is through applying "quantitative analysis techniques such as tables, graphs and statistics". He points out that this helps us to not only see but also describe and understand the relationships in the data.

As shown in Table 3, the data analysis process was divided in two steps. In step 3, the data was downloaded and grouped, and some basic statistics and calculations were performed on it. In step 4, additional calculations and manipulations were performed to clean the data from noise and outliers as well as run statistical tests on it.

More specifically, in step 3, when the data collection process was completed, the EDA readings were downloaded from Empatica in their raw form in Excel format. Excel was further used to combine the data and for further processing and analysis of the data. The decision was made to use Excel for convenience, ease of use and time constraints instead of an advanced program designed for analysis and Excel was more than capable of handling the required tasks.

The data processing began with first matching the participant number with the excel version they used as well as the time it took for the participant to complete the experiment. After this some basic calculations followed such as the average EDA per participant, the Standard Deviation, as well as the total points they scored. These values were then displayed in a series of graphs, bar charts and box plots to allow the researchers to see the changes for each of the hypotheses. These initial analyses gave a good overview of the results and allowed to spot outliers as well as the need to clean data from noise. For example, putting the 1-2-minute baseline period in mind, the initial 100 EDA values for each participant were excluded from some of the further calculations.

Based on the support of the hypotheses, the researchers could move to step 4 to further analyse the relationships between the findings in additional ways. This is expected to be done using the Data Analysis tool pack in Excel. The data is grouped as necessary to perform the statistical testing if required.

3.8 Validity and Reliability

In general validity refers to the strength of the conclusion in a research paper. It applies to both method and design. It can refer to the internal validity, which is the affected by the flaw in the study itself, or the external validity, which is how the sample of the research can be compared to the entire population (Saunders 2016). The research philosophy of post positivism (or critical realism) suggests that the one reality is hard to grasp and many aspects have to be considered. The experiment itself was conducted in a controlled environment, but the day of the participants before the test wasn't. EDA measurement is a complex process in itself, with different variables. This setting provided the opportunity test real life experience and not completely laboratory settings, which paints a picture closer to that one reality. External validity was a variable considered in sampling and the research is showing a broad picture in novice decision making. The sample size also strengthened the internal validity through in-depth analysis and control, which enabled to exclude outlier values.

Reliability is defined as the consistency and repeatability of the measurement (Saunders 2016). In terms of EDA measurement, it is likely that the individual values would be different, but it is according to the structure of EDA. With the presented indicators in Findings, these different measures can be compensated, as the trends and changes are the interesting aspect of EDA measurement, not the actual values. These precautions made the research reliable and able to be repeated in a proper manner with the same outcome.

4. Findings

This chapter presents the results of the conducted experiment, as well as the analysis and testing of these results. It further outlines the errors that were encountered during the experiment process and how they were solved. First, the outputs are each described separately to provide a general overview of the gathered data, putting a high emphasis on the EDA readings. Then the calculated indicators, which were used throughout the various analysis steps during the research, are presented along with an explanation of why they were chosen and their implications. Next, the framework used to evaluate decision quality and how it can be transformed into data with statistical value, is laid out. Followed by this, the process of testing the hypotheses individually using the calculated indicators is presented. Finally, more in-depth statistical manipulations are presented to test the relationships and interconnections between the variables are presented.

4.1 Descriptive Statistics

4.1.1 The measurement process

The experiment process began after the experiment design was tested on two volunteers, who were later excluded from live versions testing to keep consistency. The testing period lasted approximately for two weeks, during which, 33 tests were conducted through the protocol discussed in the Methodology. The number of tests derives from the number of sub-hypotheses related to the dataset, which is four. It generated 16 Excel versions, so each version was used in the test exactly two times, except for one, which had to be redone, due to human errors, which are discussed in subsequent sections.

It is important to lay down some ground rules for the measurement. Tonic EDA (SCL) is an ever-changing baseline. The value is constantly changing with the individual and can be vastly different between two persons. As SCR is also included in the measurement, it can provide inaccurate information in terms of SCL measurement (Braithwaite et. al. 2013). As there are many determining factors which can change the outcome of the measurement, it is crucial to strive to provide the same exact environment for all the participants during the test. Many of these elements are beyond our control, such as the weather or the participants' mood before the experiment; while some are limited by the boundaries of the research, for example the time constraints or the way volunteers can be gathered. Other than the factors beyond our control, the tests were done in a controlled environment. The uncontrolled elements, on the other hand, provide a more naturalistic view. The participants were greeted, then shortly introduced to the task and the EDA measuring wristband. It was explained, that the wristband works in a similar way and provides a similar feeling as a regular smart watch. Every volunteer was told, that they have roughly 30 minutes to complete the test. Fixing the length of the test would have been beneficial to have equal measure length, just as the literature suggests, it, however, was not feasible. As expected, some participants completed the task highly ahead of time, while others several exceptions took double the time. This created another dimension, along which, data can be analysed. The participants had to work on a Windows based laptop, provided by the researchers. They were shown the Google Form which contained a short description for the case, as well as the Excel dataset to base their answers on. The EDA measurement started at the same time as the screen capture, which provided a solution to detect the events that resulted in EDA spikes and drops during measurement. Both researchers were present during all of the tests, which were held in previously booked empty conference rooms at Copenhagen Business School, which provided optimal physical conditions for the test. The participants were told that if they had trouble understanding the questions, not the data or dataset, they were free to ask. When they completed the task, the screen recording, and the EDA data stream were stopped at the same time. Upon completion the participants were thanked and given some chocolate snacks; this was followed by a short explanation as to what the goal of the experiment was and if they wanted to, they were shown their EDA graphs.

The experiment with the experts was conducted in the same way, with only two exceptions. The first exception being a different location, which was a conference room at their office building. The second, exception was that EDA measurement was optional for them. A total of five experts agreed to participate in the research, and their inputs were converted into a scoring framework to determine the decision quality of the novice participants.

4.1.2 Problems encountered

The first test provided an opportunity to encounter many of the possible human mistakes. This measurement was inaccurate for two reasons; first, the charging dock of the wristband was left on the watch, meaning the EDA sensor was not in contact with the skin of the participant. Secondly, due to a weak internet signal in the room the survey had to be refreshed and the answers filled out again, so even if the measurement was correct it would be difficult to track the triggers in EDA changes.

The E4 wristband, supplied by Copenhagen Business School, provided a great opportunity to delve into something new and exciting as well as broaden the research opportunities. Simultaneously it gave the researchers an understanding that certain scientific research tools are not perfect. The available measurements were considered carefully. The wristband provides several measures of physiological attributes, but during the equipment testing step, it was discovered that only the EDA measurements were accurate. The heart rate measurement was also considered for analysis, but due to inaccuracies the decision was made not to us it. In order to check the credibility of the EDA measurement, a benchmarking test was conducted with two different wristbands on both hands of one of the researchers. The base values were slightly different, as expected due to the nature and composition of EDA, but the overall changes and spikes were consistent on both devices.

During the experiment several problems with the EDA measurement were encountered. The data streaming is Bluetooth based, which had a weak signal on one occasion for a few minutes but reconnected shortly after and as the data output is in one excel column, the combination of two separate measurement provided no problem. As the readings were consistent enough around the disconnection, it was considered as a valid reading. The EDA reading occasionally dropped to a value of zero, so cleaning of the data was necessary. Normalization is defined as necessary data transformation for the data in order to provide basis for credible statistical analysis to filter out skew and kurtosis (Braithwaite, 2013). This process does not help with the between-participant comparisons. On the other hand, standardization refers to the data cleaning process, which enables direct and meaningful comparison between the volunteers (Braithwaite, 2013). These processes will be further discussed at the introduction of the collected and aggregated EDA readings.

4.1.3 Data collection overview

Due to the complexity of the test, several inputs were required, and generated several different data outputs which can be used for analysis. The most important data collected was the 33 EDA measurements recorded with the wristband. Apart from case questions, the Google Form provided data on the participants' level of data analysis experience and Excel proficiency (the tool used in the experiment), some demographic data on gender and age of the participants, as well as the perceived cognitive load and overall mood before and after solving the case. The screen recording is considered as another layer of collected data, to further secure the data and in order to provide the option to perform a deeper analysis through that source, if necessary.

The sample of volunteers, as previously described, consists of friends, co-workers and fellow students, people who are at the end or close to finishing their studies and joining the professional market full time. This statement is consistent with the determined average age in the survey, as 23 persons stated that they are between 18 and 25 years old, and 10 persons stated that they are between 26 and 35 years old, which puts the average age to around 24, just below the average of age of finishing a master's program. The gender distribution is well balanced, 17 female and 16 male participants took the test, which can be considered a successful sample.

In terms of data analysis experience and Excel proficiency the answers are quite distributed amongst the lower and mid-levels, but no one qualified themselves on the highest, expert level on either of these questions. The distribution of the answers is almost identical, with Excel proficiency having a slightly higher average value of 2,91 out of 5 compared to 2,62 out of 5 on data analysis experience.

The question of overall mood was brought up before and after the case questions were completed, with the range of very calm and very overwhelmed both times. In 14, almost half the cases, this answer did not change throughout the test. 10 participants felt more overwhelmed at the end of the test, and in 8 cases the volunteers felt more relieved at the end. The expected outcome is being more frustrated at the end, but the opposite can be explained by the relief of finishing a test, that was advertised as something complicated. The average mood was 2,66 before, and 2,77 after, which is just barely on the calmer side of the spectrum.

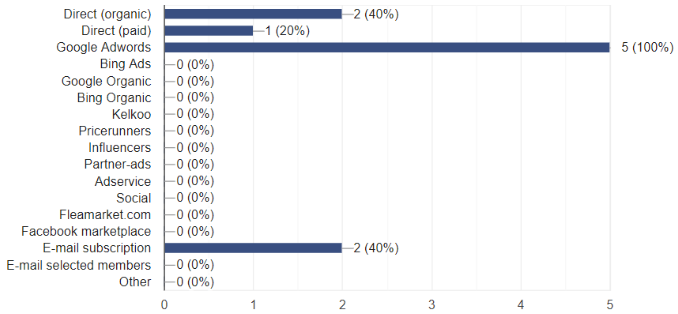
The end of the survey contained four questions related to perceived cognitive load, each focusing on different types of cognitive load (intrinsic, extraneous and germane). Overall the participants found the exercise quite difficult, as the average difficulty on a scale of 1 to 5 was 3,58. The complexity of the dataset was slightly lower, 3,06, which means the perceived difficulty derives from the nature of the questions, which is understandable as everyone was aware that there aren't universally agreed good answers as the data is missing various information dimensions, such as budget. This presumption is further confirmed by the fact that the difficulty of the exercise was also considered lower than the difficulty of the task, with an average of 3,16. The last question was about the perceived level of focus during the test, which was expected to be around the average mood after the test, but the level of focus proved to be higher, 3,63, the highest average found on the 1-5 scale. It can be described by looking at the consequences of the test. The participants were familiar with the fact that there are no straight good answers to the questions, and they also don't get graded, based on how they do, and neither do they get the point received by the method how decision quality is calculated.

4.1.4 Pointing framework and case questions

As previously mentioned in the Methodology, the decision quality is determined by comparing the novice answers to the answers by the experts. Five experts participated in the test, which means 9 answers, as 2 answers were expected by them as well for every question, except question 3 about the platforms. They were provided the exact same Google form, with a different copy made to see their answers separately. The answers are available below in Figure 5.

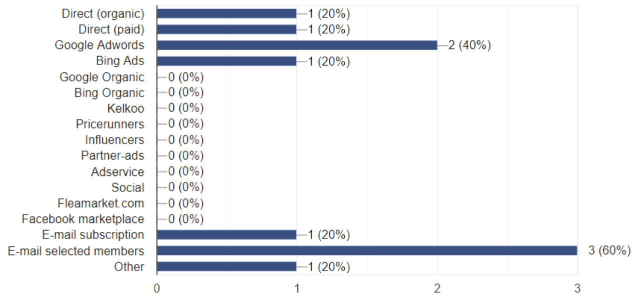
1. Which two channels do you think are the most successful overall? Please choose 2 channels.

5 responses



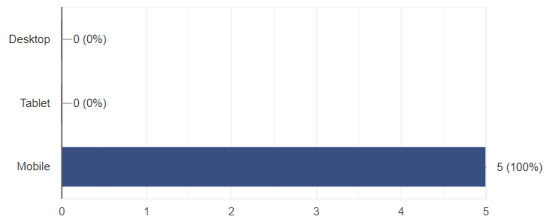
2. Some channels are reaching their potential. Which of these channels are the most saturated? Please choose 2 channels.

5 responses



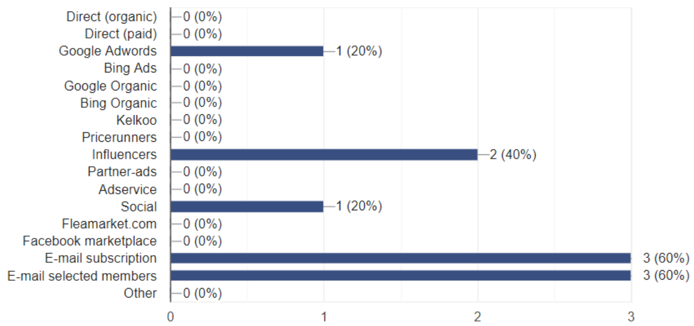
3. How do you expect the platform distribution will change in a year from now? Strategically, which platform would you focus on developing? (Desktop-Tablet-Mobile) Please choose a platform.

5 responses



4. Which channel would you choose as the most suitable for a special promotion to boost short term visits? Please choose 2 channels.

5 responses



5. Based on the information in sheet 4 (Responsibilities), to which channel(s) would you allocate additional work hours? Please choose 2 channels.

5 responses

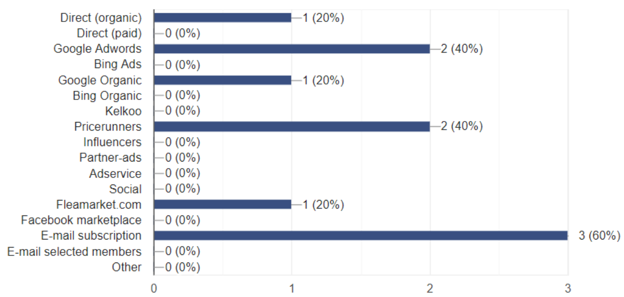


Figure 5: Expert Survey Responses

Different tiers of answers were created based on the frequency of how many experts choose that option. Three tiers were created, where tier three is the most frequent answer and is worth 3 points, tier two answers worth 2, and tier one answers are worth 1 point. The actual answers and the framework of the point system is available in Table 4 below. As we can see, the theoretical maximum point of 27 (based on 9 data entries) is not available, as for the double answer questions, only one had two tier two answers, which puts the maximum reachable points to 24 (5,5,3,6,5 respectively in the questions order).

Points	Question 1	Question 2	Question 3	Question 4	Question 5
3 points	Google AdWords	E-mail selected members	Mobile	E-mail subscription	E-mail subscription
	-	-	-	E-mail selected members	-
2 points	E-mail subscription	Google AdWords	-	Influencers	Google AdWords
	Direct (organic)	-	-	-	Pricerunner
1 point	Direct (paid)	Direct (organic)	-	Google AdWords	Direct (organic)
	-	Direct (paid)	-	Social	Fleamarket.com
	-	Bing Ads	-	-	-
	-	E-mail subscription	-	-	-

Table 4: Point Distribution Framework used to determine the performance of the participants

No one was able to reach the available maximum points of 24. The highest achieved point was 18, which was earned by two different participants. The average of the points was 11,25. The point averages per question were 2,86; 1,63; 2,34; 2,69 and 1,72 respectively. Keep in mind, the maximum available points were 5; 5; 3; 6 and 5. Based on the tests, the outcome of question 2 is the lowest as expected, almost no one could interpret the question on their own, as the participants were not familiar with the meaning of channel saturation. The low average of question 5 was also expected as it had the most complaints that is too subjective and the information dimensions are too restricted. Question 3 was surprising, 7 people choose Desktop as the platform to innovate on, even though Mobile is widely considered the dominant platform in almost any possible aspect. This idea was backed up by the expert's opinion, as all of them without an exception has selected Mobile. The other two questions were answered according to expectations.

4.1.6 Time as a factor

In terms of timing, as previously mentioned, the participants were told they have 20-30 minutes available, but also that the test wouldn't be stopped at any point, as answers were necessary for further evaluation, and it provides other means of analysing the data as time is considered a factor for decision quality shown on the initial model. The average time to complete the test was just barely above 30 minutes, but it can be interesting to look at the standard deviation as some measurements were surprisingly low and other tests took a long time to complete. The shortest test took 8 minutes and 13 seconds (rounded to 8 minutes), while the longest was 1 hour 16 minutes and 2 seconds (rounded to 76 minutes), while the whole sample had a standard deviation of 14,85 measured in minutes for comprehensible numbers, which is almost double the duration of the shortest test.

Two types of the data points were not mentioned in this section before. One of them is the second part of each question, where participants were expected to explain their choices of channels. This generated a large volume of text, which was necessary to create an immersion of a serious test, where random guessing in contrast of educated guessing was discouraged, and the latter was expected. This collected data absolutely falls out of the scope of this research but skimming through the answers provides some idea what kind of data inspired some of the decisions for example providing the specific metric in the explanation such as "Highest average transactions" or "Conversion rate vs. traffic". On the other hand, in some cases it further strengthens the idea that some answers are just educated guesses. Handful of examples are available for this situation such as "This is just a guess" or "I have no idea".

The other unexplained dataset is the screen recording. As previously mentioned OBS (Open Broadcaster Software) was used to capture the laptop screens while the participants took the tests on it, with the intent to provide a way to properly time the events seen on the EDA measurements, as the application for the wristband doesn't provide such option. The screen capture started at the same time as the EDA measurement, and the overall measured time is consistent with the EDA readings at 961 minutes. This extra step further helped to examine which datasheet was used to answer each question. A surprising find is that even though the "2019 Averages" sheet provided the least amount of pure information and was more of a sheet that provides condensed data, it was used by almost every participant with high frequency. Overall the data analysis was more focused on descriptive and statistical analysis, which meant the screen capture provides less value than expected.

4.1.7 EDA readings

The most crucial data, which enabled this research was the EDA measured with galvanic skin response wristband. It was measured for all 33 participants through two different wristbands. After account creation, and logging into the smartphone application the following user interface becomes available which we can be seen on the left side of Figure 6.

This application enables to connect to the devices through Bluetooth. Figure 6 shows the interface after the data starts streaming, where the various graphs help to follow the real-time streamed data. During the tests, the stream was monitored to ensure the validity of the collected data.

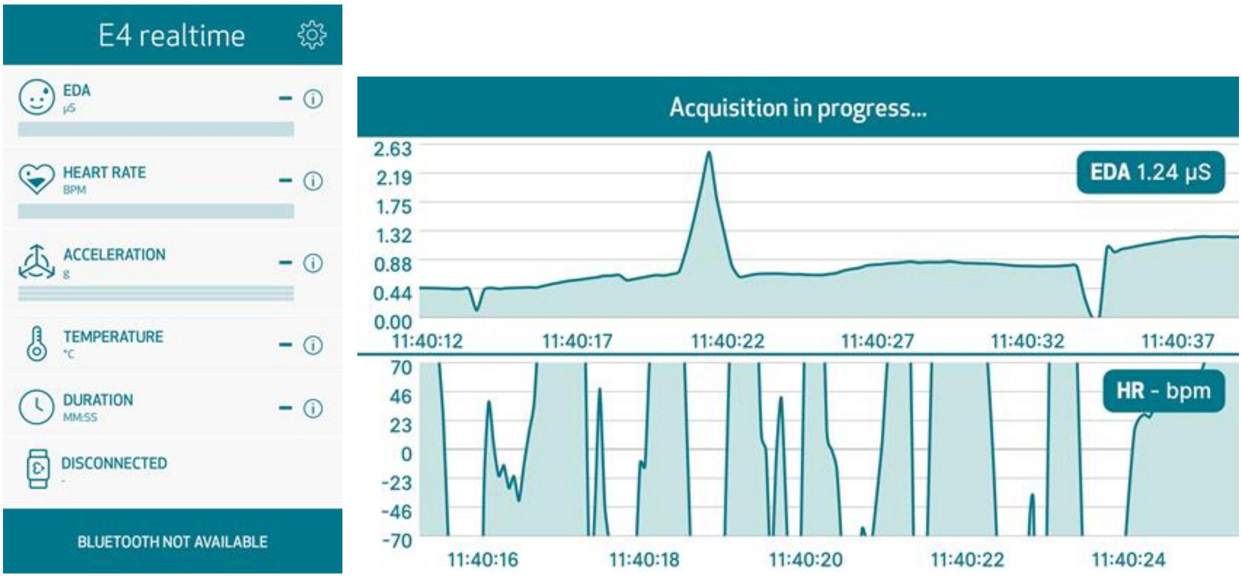


Figure 6: Application Interface and Streaming Interface

The outcome of the measurement is stored on the webpage (Empatica, 2019) and is available after logging in (presented on Figure 6). The first graph in blue shows the EDA reading, which the research is interested in. The following graphs are Blood Volume Pulse, Acceleration, Heart Rate and Temperature respectively. They are not considered for analysis, as it is visible from the measurements such as the numbers for temperature or for example on Figure 6, they are not reliable.

Looking at the data, and more specifically the graphs, two aspects were evident. As talked about it previously, EDA is built up from different components changing from person to person, also determined by the situation, makes it hard to measure. Outstandingly high and surprisingly low values were both found, with the lowest value being below 0,02 and the highest being just a bit shy of 18. The other aspect to look at is the trend of change during the test. Different variations have occurred ranging from slight increase all throughout the test, to steadily decreasing. The frequency of change is also inconsistent as some readings show a nice graph of steady change in terms of increase or decrease while other measurements are more chaotic with several local maximum and minimum values. As these readings are so variated, a more in-depth analysis is required to evaluate the data, as expected.

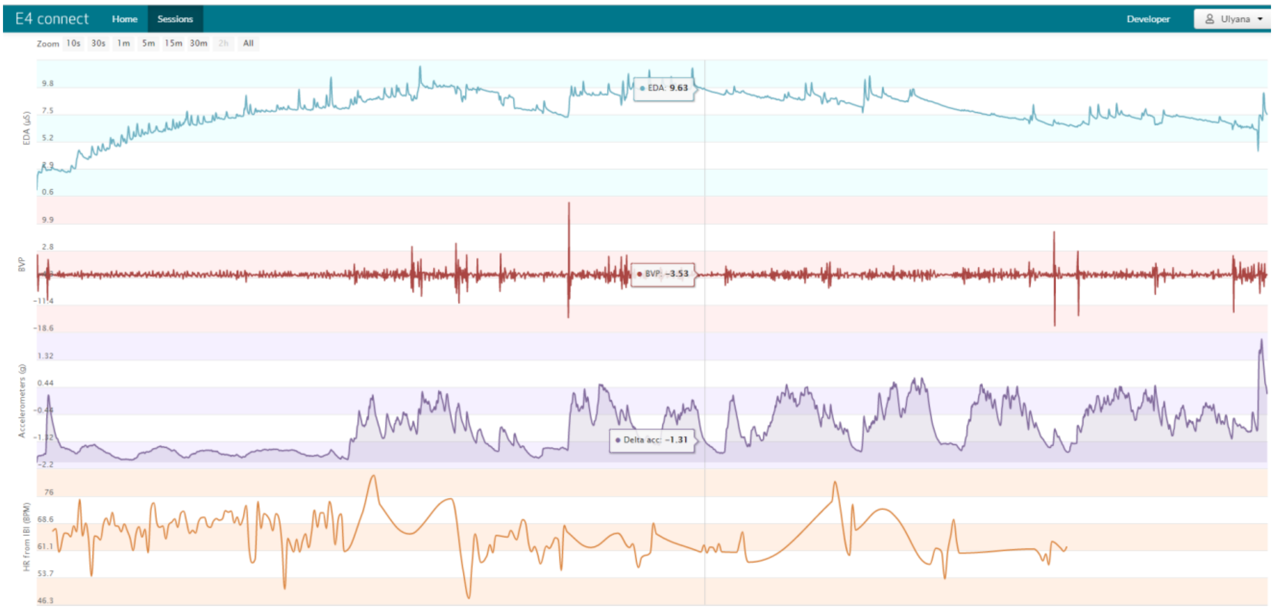


Figure 7: Example of a final reading from a participant

4.1.8 Aggregated data

The data generated by the EDA measurement also comes in an Excel format, which made it possible to collect it in one big Excel file where data could be aggregated and put into use for descriptive statistical analysis seen in the table below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	**Participant number**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
2	**Excel Version**	***0010**	***0100**	***0000**	***0001**	***0110**	***1000**	***0101**	***1011**	***1001**	***1010**	***0111**	***0011**	***1100**	***1111**	***1101**
3	**Duration**	0h27m42s	0h49m43s	0h35m13s	0h28h44s	0h20m36s	0h19m5s	0h22m33s	0h38m3s	0h21m43s	0h30m57s	0h25m52s	0h14m1s	0h16m42s	0h8m13s	0h32m7s
4	**Duration (mins rounded)**	28	50	35	29	21	19	23	38	22	31	26	14	17	9	32
5	**Points Scored**	10	15	12	8	13	9	7	11	13	17	7	15	5	5	13
6	**Points per minute**	0.36	0.30	0.34	0.28	0.62	0.47	0.30	0.29	0.59	0.55	0.27	1.07	0.29	0.56	0.41
7	**Mean EDA**	0.15	4.70	2.57	0.38	0.12	10.57	2.80	11.23	0.27	0.41	0.54	0.23	8.13	0.45	9.43
8	**Standard Deviation**	0.06	1.15	0.46	0.25	0.11	1.36	0.52	3.87	0.01	0.05	0.28	0.07	1.32	0.14	2.80
9	**Variance**	0.00	1.32	0.21	0.06	0.01	1.85	0.27	15.01	0.00	0.00	0.08	0.01	1.75	0.02	7.84
10	**MIN EDA**	0.01	1.14	1.66	0.15	0.08	7.38	1.53	4.91	0.20	0.28	0.19	0.14	5.75	0.28	2.55
11	**MAX EDA**	0.45	7.19	3.57	2.44	0.66	12.31	4.13	17.84	0.31	0.83	2.11	0.48	11.35	2.55	14.83
12	**Maximum distance**	0.44	6.05	1.92	2.29	0.59	4.93	2.60	12.93	0.11	0.55	1.92	0.34	5.60	2.26	12.27
13	**Maximum distance: relative**	4412.48%	629.98%	215.69%	1590.41%	876.27%	166.89%	269.66%	363.63%	152.50%	300.27%	1117.80%	351.40%	197.44%	899.32%	580.57%
14	**MAX Average EDA**	0.30	2.49	1.00	2.07	0.54	1.74	1.33	6.61	0.04	0.41	1.57	0.26	3.22	2.09	5.39
15		297%	153%	139%	651%	554%	116%	148%	159%	116%	200%	388%	214%	140%	561%	157%
16																
17	**Average EDA: Start**	0.256	1.247	2.406	0.162	0.529	7.654	1.675	9.991	0.232	0.381	1.007	0.413	8.099	0.312	3.211
18	**SD: Start**	0.010	0.087	0.044	0.005	0.025	0.157	0.051	0.135	0.006	0.018	0.066	0.024	0.368	0.024	0.318
19	**Average EDA: Final**	0.087	5.758	1.705	1.064	0.085	11.473	3.388	5.475	0.274	0.289	1.688	0.142	6.367	0.777	10.673
20	**SD: Final**	0.027	0.431	0.041	0.068	0.002	0.155	0.309	0.076	0.003	0.013	0.190	0.001	0.130	0.154	0.212
21	**EDA Change**	0.34	4.62	0.71	6.57	0.16	1.50	2.02	0.55	1.18	0.76	1.68	0.34	0.79	2.49	3.32
22	**LQ Start**	0.247	1.169	2.380	0.156	0.511	7.503	1.636	9.962	0.231	0.369	0.947	0.400	7.858	0.293	3.049
23	**UQ Start**	0.265	1.314	2.432	0.165	0.540	7.783	1.704	9.995	0.236	0.393	1.055	0.419	8.298	0.317	3.498
24		0.256	1.241	2.406	0.161	0.525	7.643	1.670	9.979	0.233	0.381	1.001	0.409	8.078	0.305	3.274
25	**LQ Final**	0.0922	5.8569	1.67	1.01	0.08	11.37	3.08	5.42	0.27	0.28	1.59	0.14	6.34	0.67	10.48
26	**UQ Final**	0.0948	5.9274	1.74	1.11	0.09	11.51	3.64	5.54	0.27	0.29	1.80	0.14	6.40	0.94	10.88
27		0.0935	5.8922	1.7042	1.0590	0.0852	11.4409	3.3604	5.4799	0.2741	0.2866	1.6926	0.1422	6.3674	0.8053	10.6819
28	**Clean EDA Change**	0.36	4.75	0.71	6.59	0.16	1.50	2.01	0.55	1.18	0.75	1.69	0.35	0.79	2.64	3.26
29																
30																
31	**EDA**	0.174198	0.837688	0.837688	0.111436	0.429091	0.837688	0.836457	0.837688	0.125525	0.288213	0.837688	0.220309	0.837688	0.178041	0.837688
32		0.249769	1.18507	1.18507	0.152423	0.535403	1.18507	1.18386	1.18507	0.199815	0.356103	1.18507	0.297161	1.18507	0.26514	1.18507
33		0.270263	0.936651	1.623196	0.158828	0.56102	1.623196	1.26591	1.623196	0.208782	0.356103	1.045524	0.313813	1.623196	0.281791	1.623196
34		0.281791	0.954583	1.983219	0.163951	0.573829	2.101059	1.3069	2.101059	0.213905	0.361227	1.096759	0.325341	2.101059	0.290757	1.914052
35		0.286914	0.967391	2.00007	0.166513	0.580233	2.589269	1.331238	2.589269	0.216467	0.361227	1.130062	0.329183	2.589269	0.295881	1.943512
36		0.292038	0.972515	2.021844	0.166513	0.582795	3.110874	1.345328	3.110874	0.217748	0.358665	1.145432	0.333026	3.110874	0.297161	1.964205

Figure 8: A partial capture of the final Collected Data-Set

The first column contains the labels, starting with the previously described information of each test, the previous characteristics of the data, which were mentioned and at the bottom below "EDA measurement" cell, we can see the actual measurement from the wristband. The Excel is close to 20, 000 rows, as one second of wristband measurement generates between 3 and 5 datapoints. The columns contain each participant separately, first providing their ID and the Excel version they used for the test. This is going to be used to group them based on the hypotheses for further calculations on other sheets. Next rows contain the exact duration, and a version that has been rounded to minutes to make it easier to comprehend the data. Points are included for the participants, calculated based on the decision quality framework and the answers found in the Google form for the case questions. The points were distributed with the time it took for each participant to complete the test, defining their points/minutes score, which puts us closer to compare decision quality between participants.

The next section contains the basic statistical calculations, which unveils various statistical indicators used for further analysis. First the mean EDA was calculated throughout all the data points. As it is visible from the difference in values, EDA is built up by components and is different for everyone, this average is not sufficient to draw conclusions from, especially with a sample size like this. The next variable to calculate was the standard deviation, to determine how consistent are these values. As previously mentioned, the graphs had several variations in trends and consistency, so the standard deviation also showed values in a wide interval. The next indicator to calculate was the range and relative range in values. In order to calculate that, Minimum and Maximum values had to be calculated. Max values provided no problem, but the Minimum values were exactly the same in many cases, which was probably caused by the start-up measurement of the device. In order to provide a credible minimum value, the first hundred values were excluded from this calculation. These enabled the calculation to determine the range and the relative range. These values also proved to have a higher number of outliers than preferable.

The difference between the starting and ending value was also considered valuable. In order to calculate this, the averages of the second 100 values were taken (still excluding the first 100 values for minimum) and was compared to the average of the last 100 values. It was necessary in order to avoid further outliers, as values can be really different even in a span of a second. EDA Change was calculated by distributing this calculated final value with the starting one. It produced a more digestible indicator ranging from the value of 0,16 to 6,57. In order to further exclude outlier values, the first and third quartile were determined in the previous two samples of 100 values and data below the first quartile and data above the third quartile were excluded. The difference of the third and first quartile was calculated in both cases, and Clean EDA Change was defined as the division of the difference between quartiles in the final with the difference of quartiles in the starting.

The reason that these many variables are necessary is because the sample size was chosen to be low, which on the other hand, enables us to go more in-depth on the analysis. It provides a way to filter out noise and to filter out outliers. In an ideal world all test would show a similar result, but as we have a post positivist philosophy, the one reality is hard to grasp.

Theses calculation concludes the descriptive part of the findings. In the next part, these numbers will be tested through statistical methods to evaluate them grouped by hypotheses to determine which can be supported and which can be dismissed.

4.1.9 General Overview of the Data

To begin with, several graphs were made to compare the various EDA values against the points scored and points scored per minute. Figure 9 below presents an example of one of these graphs whereby Clean EDA Change is plotted with the Points Scored per minute. Some interesting observations can be made from the graph; for example, comparing Clean EDA Change of the top two scorers (who both scored 18 points), participant 22 has a Clean EDA Change score of 1.74 lower than participant 24. At the same time, comparing Clean EDA Change of the two bottom scorers (who both scored 5 points), participant 13's Clean EDA Change score is 1.85 lower than that of participant 14. Moreover, participant 22 and participant 13 have a difference of 13 total points scored while their Clean EDA Change scores are 0.55 and 0.79 respectively. This is only a 0.24 difference in stress level but such a drastic difference in points.

Additionally, participants 4 and 25 are interesting to compare. While they both scored the same amount of points (8 points) participant 4 had the highest Clean EDA Change score of 6.59 among all participants. This score is 6.03 higher than that of participant 25, and 4.91 higher than the average Clean EDA Change score.

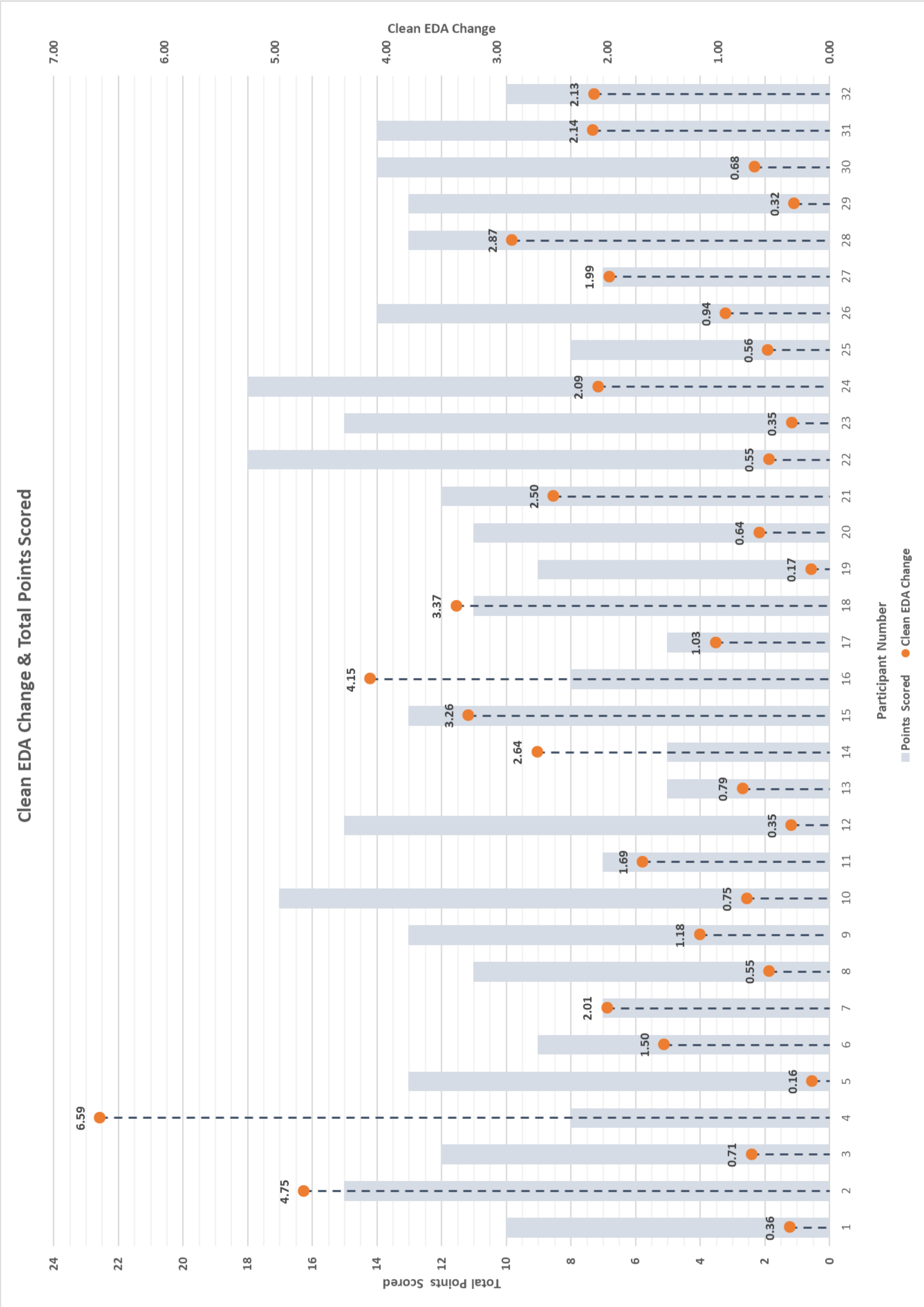


Figure 9: Clean EDA Change plotted against the Total Points Scored by each participant

4.2 Hypothesis Testing

Basic descriptive statistics provide a general overview of the data, but they do not allow for a better, deeper understanding of the data and the significant results. Therefore, several statistical measures were calculated to test whether the predictions made in each hypothesis were supported or not. The following section describes and explains the first steps of testing the hypotheses of this study.

After the data was collected and combined, and analysed using basic calculations, the hypotheses were tested using additional statistical calculations and manipulations. For hypotheses 2-5 these calculations, for example, included the mean EDA of each participant and the standard deviation (which will be referred to as SD from now on) of the EDA of each participant. Since hypothesis 1 differs from hypotheses 2-5 additional calculations were performed, such as the total points scored as well as the points scored per measure of time (in this case per minute minute). However, as the mean EDA and SD were based on pure raw data the decision was made to clear the noise and outliers resulting in two additional values EDA Change and Clean EDA Change. These results are further presented individually in the sections below.

Additionally, for hypotheses 2 to 5, all the values were grouped such that there were two groups of 16 values per hypothesis. An example of this distribution for mean EDA can be observed in Figure 10; for each of the four hypotheses the group where the variable was absent is denoted as the control group and where the variable was present it is denoted as the test group. Grouping data in this way gives a clearer view of the effect of the variables on the cognitive load of the participants.

Hypothesis 2a		Hypothesis 3a		Hypothesis 4a		Hypothesis 5a	
control	test	control	test	control	test	control	test
0.1523	11.2259	0.1523	4.7036	4.7036	0.1523	0.1523	0.3757
4.7036	0.2706	2.5717	0.1196	2.5717	0.1196	4.7036	2.7981
2.5717	0.4135	0.3757	2.7981	0.3757	11.2259	2.5717	11.2259
0.3757	8.1333	10.5729	0.5426	10.5729	0.4135	0.1196	0.2706
0.1196	0.4539	11.2259	8.1333	2.7981	0.5426	10.5729	0.5426
2.7981	9.4331	0.2706	0.4539	0.2706	0.2253	0.4135	0.2253
0.5426	1.1906	0.4135	9.4331	8.1333	0.4539	8.1333	0.4539
0.2253	1.0396	0.2253	1.1906	9.4331	1.1906	1.1906	9.4331
0.2177	7.8870	0.1535	0.2177	0.2177	0.1535	0.2177	1.0396
0.1535	5.0543	1.0396	0.1651	1.0396	0.1651	0.1535	7.8870
0.1651	0.1339	7.8870	5.0543	5.0543	7.8870	0.1651	0.9156
0.9156	3.7471	0.9156	0.9345	0.9156	0.1339	5.0543	0.7545
0.7545	3.0334	0.1339	0.1388	0.9345	0.7545	0.1339	0.9345
0.9345	10.5729	0.7545	3.7471	3.7471	0.1388	0.1250	0.1388
0.1388	8.6512	0.1250	3.0334	0.1250	3.0334	3.0334	3.7471
0.1250	0.2624	0.2624	8.6512	0.2624	8.6512	0.2624	8.6512

Figure 10: Mean EDA grouped by control and test groups for each hypothesis

4.2.1 Hypothesis 2a

H_{2a}: Including raw data increases cognitive load

The process of testing the hypothesis 2-5was standard for these hypotheses and began with comparing the mean EDA and SD between the groups. To help more easily compare these values they were presented using bar charts with error bars in Microsoft Excel. Figure 11 below shows this bar chart for hypothesis 2a on the left. The difference in means between the groups is clearly visible, in fact the mean EDA for the test group is 3.54 higher than for the control group. If we recall hypothesis 2a: “including raw data increases cognitive load", this suggests that when raw data was present in the dataset the participants’ stress level increased and therefore their cognitive load did too. Since the mean EDA is based on the pure raw data, additional measures were calculated namely EDA Change and Clean EDA Change (ones that have been cleaned from noise and outliers and used to create two additional bar charts. The bar chart with EDA Change and Clean EDA Change and their standard deviations can be observed below in Figure 11 in the middle and on the right respectively. Again, although a smaller one, an increase of 0.12 is evident in the test group for EDA Change and an increase of 0.13 for Clean EDA Change, suggesting that the hypothesis is likely to be true.

Moreover, +/- 1 SD was applied to both charts. It can be observed that the SD error bars overlap on all three charts in Figure 11. This means that 68% of the data for Hypothesis 1 using all three measures overlaps, thus suggesting that the increases between the two groups are not significant. Further statistical tests are therefore required to evaluate the significance of the results. Although, inferential statistics show some increase in EDA for hypothesis 2a the statistical significance of this data is tested in the next section. As the body response in general tends to be mild, these results are addressed and discussed further using relevant existing findings and theories.

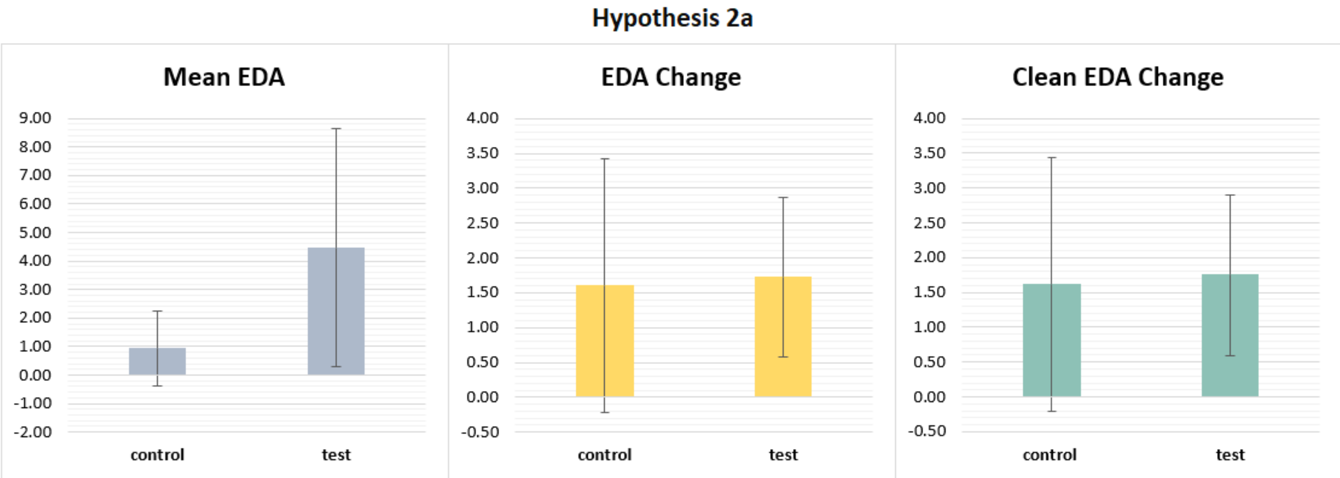


Figure 11: Box plots of mean EDA (left), EDA Change (middle) and Clean EDA Change (right) with their respective +/- 1 S.D. for Hypothesis 2a

4.2.2 Hypothesis 3a

H_{3a}: Adding new visualisation decisions increases cognitive load

Just like for other hypotheses, three values for EDA were computed: mean EDA, EDA Change and Clean EDA Change. The rest of the hypotheses will be explained focusing on EDA Change and Clean EDA Change, as they appear to be more accurate due to absence of noise and outliers. Mean EDA will still be presented for comparison. Observing the change in EDA from Figure 12 below, increases of 0.37 and 0.43 are evident between the test and control groups (middle and right charts respectively). Again, recalling the claim made in hypothesis 3a: “adding new visualisation decisions increases cognitive load”, these results suggest that stimulating the participants to make new visualisation decisions did in fact increase their stress level and highly likely their cognitive load. Interestingly this is a higher increase in stress level compared to hypothesis 2a.

Looking at the SD for EDA Change and Clean EDA Change it can be observed that they almost fully overlap on all three charts. Again, this means that 68% of the data for Hypothesis 2 using all three measures overlaps, thus suggesting that the increases between the two groups are not significant and statistical tests are required to evaluate the significance of the results.

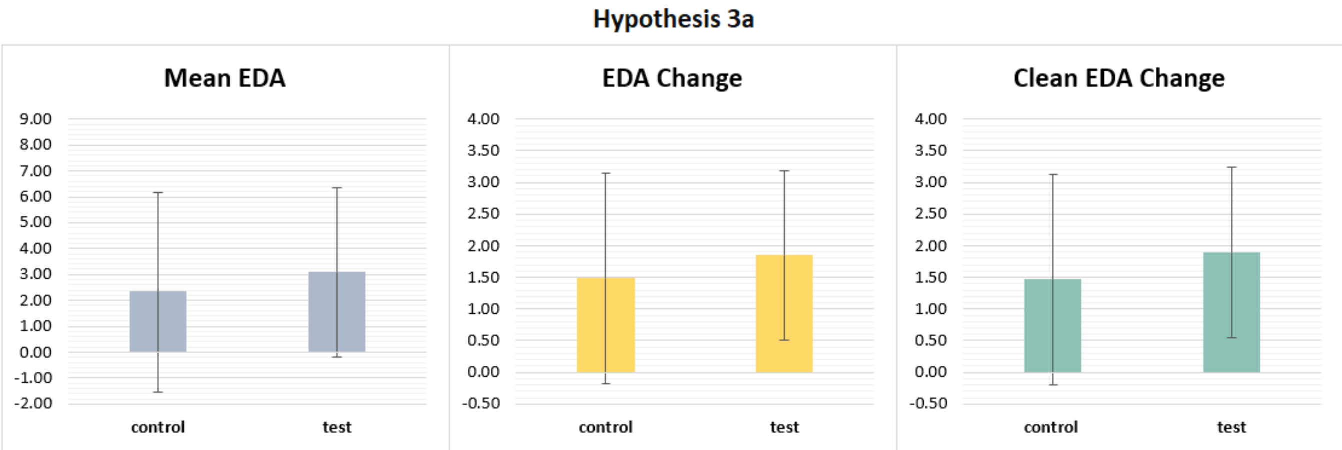


Figure 12: Box plots of mean EDA (left), EDA Change (middle) and Clean EDA Change (right) with their respective +/- 1 S.D. for Hypothesis 3a

4.2.3 Hypothesis 4a

H_{4a}: Multiple data visualisation formats increases cognitive load

With regard to hypothesis 4a, a decrease of 0.27 and 0.28 can be observed for EDA Change and Clean EDA Change values respectively in Figure 13 below (middle and right charts). Recalling the prediction made in hypothesis 4a: "multiple data visualisation formats increase cognitive load", from the results it is evident that in fact there is an opposite effect; their stress levels decrease and highly likely their cognitive load decreased as well. According to these findings, presenting data in multiple formats at the same time reduces cognitive load. This result is interestingly opposite from the results of the previous two hypotheses 2a and 3a.

Once again, looking at the SD on all three charts in Figure 13 it can be observed that it overlaps in all cases, which suggests that decreases between the two groups may not be significant and further statistical tests are required to evaluate the significance of the results.

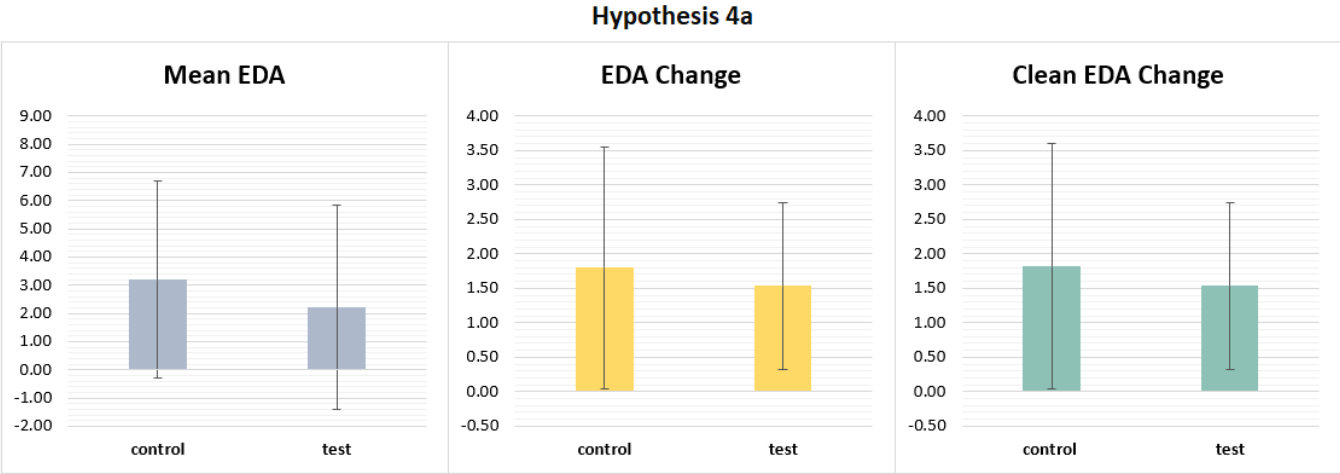


Figure 13: Box plots of mean EDA (left), EDA Change (middle) and Clean EDA Change (right) with their respective +/- 1 S.D. for Hypothesis 4a

4.2.4 Hypothesis 5a

H_{5a}: Increasing the number of colours in a data-set increases cognitive load

Finally, looking at hypothesis 5a, similarly to hypotheses 2a and 3a, increases of 0.34 and 0.36 can be observed for EDA Change and Clean EDA Change respectively. Recalling the prediction stated in hypothesis 5a: "higher number of colours in a data-set increases cognitive load", it is evident that indeed the test group that was presented a data-set full of colours experienced higher stress levels, and highly likely higher cognitive load levels as well, compared to the control group who were presented a very limited amount of colours. Comparing the SD for all three charts in Figure 14 it is evident that the SD overlaps on all three of them. Again, this suggests that 68% of the data in hypothesis 5a overlaps suggesting that the increase in EDA may not be significant and further tests are required.

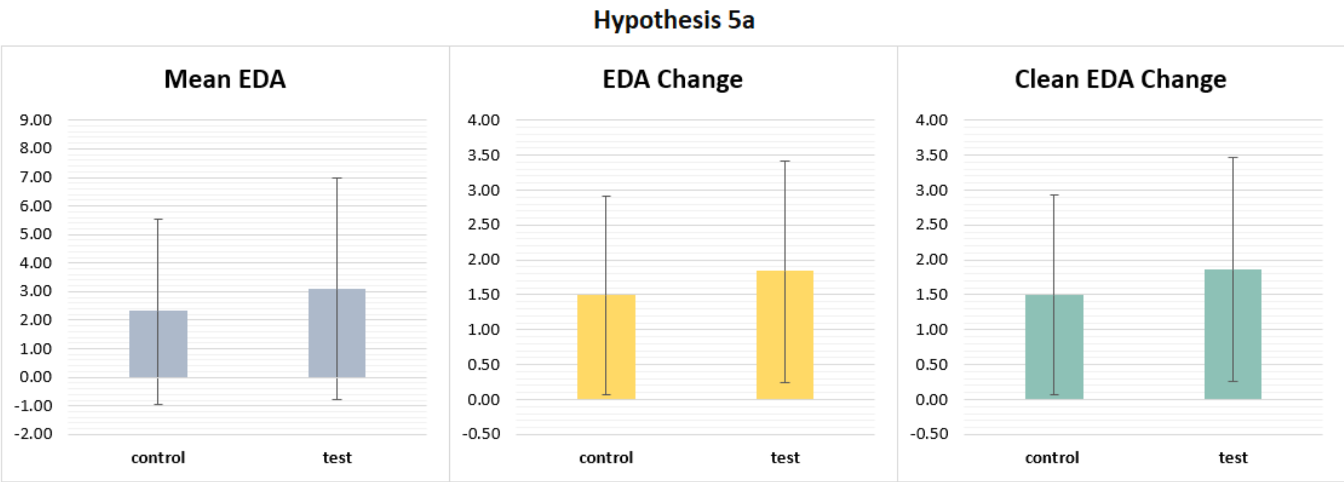


Figure 14: Box plots of mean EDA (left), EDA Change (middle) and Clean EDA Change (right) with their respective +/- 1 S.D. for Hypothesis 5a

4.2.5 Hypotheses 2b - 5b

Hypotheses 2b to 5b were the second part of hypotheses 2a to 5a respectively. The prediction in them was that presenting raw data, adding visualisation decisions, visualisation formats and excessive colours would not only increase cognitive load but also decrease decision quality. In order to see the effect these factors had on decision quality the total points scored by the participants were divided into groups. More specifically they were divided into control and test group for each hypothesis, as shown in Figure 10 above. The mean and standard deviation of these groups were then calculated and plotted as bar charts with error bars in Microsoft Excel.

Figure 15 presents the four resulting charts for each hypothesis. Observing each chart, it is evident that only Hypothesis 3b and Hypothesis 5b resulted in a decrease of decision quality (due to the lower points scored by the test group compared to control group). Interestingly, Hypothesis 2 and 4 resulted in an increase of decision quality. A deeper analysis of the relationship between EDA and the points scored (decision quality) will be described in a section below.

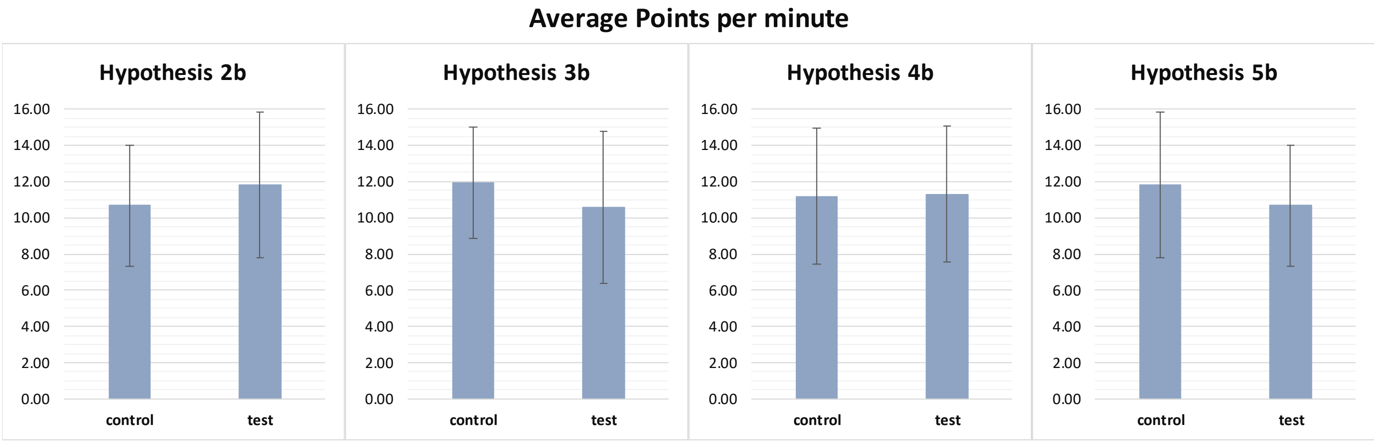


Figure 15: Average Points Scored with their respective +/-1 S.D. compared between control and tests groups for hypotheses 2b to 5b

4.2.6 Hypothesis 1

H_1: Increased cognitive load decreases decision quality

Hypothesis 1 differs from the previous four hypotheses in a way that it is the main, global hypothesis of this study that encompasses hypotheses 2 to 5. Therefore, the approach to testing it was different. The decision was made to use a combination of measures, some that were used for hypotheses 2 to 5, such as EDA Change and Clean EDA Change, and add new additional measures such as total points scored and points per measure of time. Recalling the claim made in hypothesis 1: “increased cognitive load decreases decision quality”. In this case, increased cognitive load is the independent variable, measured by EDA Change and Clean EDA Change, while decision quality is the dependent variable measured by the points the participants scored.

Figure 16 below presents a scatterplot of Points Scored per minute against Clean EDA Change. Firstly, three clear outliers can be observed in the scatterplot, that is the three participants who scored above 1,00 points per minute. The researchers attempted to remove these three values, however this did not have a significant impact on the trendline, so the decision was made to keep them. Overall, we can observe that very few values are near the trendline. However, a downward trend in the values is still noticeable. This means that the higher the change in EDA the more poorly people perform in terms of points. This therefore suggests that the prediction made in the hypothesis is supported by the data. Although, this change is minimal the researchers believe it is nonetheless valuable since the experiment conditions were kept as close to real life as possible to measure a natural organic change in EDA. This result will be further explained in the Discussion.

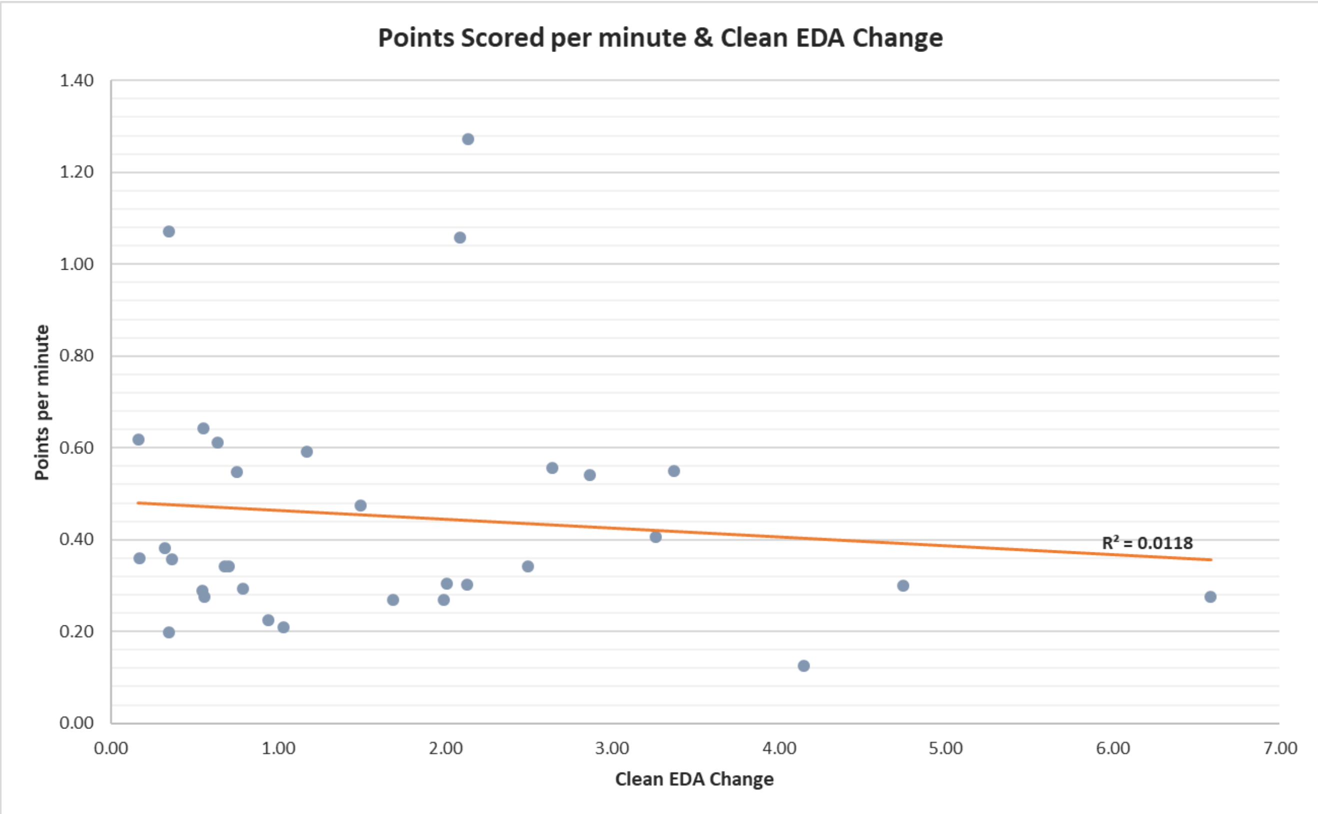


Figure 16: A scatterplot showing Points Scored per minute plotted against Clean EDA Change

4.3 Post Hoc Analysis

During the hypothesis testing step, an overlap in the standard deviation between control and test groups was observed. Therefore, the decision was made to conduct further analysis and statistical tests. Post hoc analysis refers to the statistical tests which were not specified before the data was collected. Since several new tests were conducted, the findings of these new tests had to be considered carefully, otherwise it would be considered data dredging, or in other words just trying to prove the theory trespassing the rules of the methods. More specifically, the additional tests were the two-sample t-Test and ANOVA, as well regression testing for each of the hypotheses. Moreover, a comparison was made between perceived cognitive load (recorded through survey responses) and measured cognitive load (through the EDA measurement).

4.3.1 ANOVA and t-Test

Having divided all the values (mean EDA, EDA Change and Clean EDA Change) in groups as was shown in Figure 10, Microsoft Excel was used to first run the ANOVA Single Factor test for each hypothesis followed by two-sample t-Test assuming unequal variances. An example of these two tests for Hypothesis 1 and using EDA Change values is presented in Figure 17.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
0	16	25.62133	1.601333	3.299702
1	16	27.56265	1.722665	1.29728

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.117773	1	0.117773	0.051239	0.822456	4.170877
Within Groups	68.95473	30	2.298491			
Total	69.07251	31				

t-Test: Two-Sample Assuming Unequal Variances

	0	*1*
Mean	1.601333	1.722665
Variance	3.299702	1.29728
Observations	16	16
Hypothesized Mean Difference	0	
df	25	
t Stat	-0.22636	
P(T<=t) one-tail	0.411381	
t Critical one-tail	1.708141	
P(T<=t) two-tail	0.822762	
t Critical two-tail	2.059539	

Figure 17: ANOVA: Single Factor and two-sample t-Test performed on EDA Change values

In order to evaluate the validity of the hypotheses, the values of interest in these tests were the p-values which are summarised in Table 5 below. The p-values are divided by the hypothesis, type of value used, and the test performed. They are further interpreted later in this section.

	p-values			
	mean EDA	**EDA Change**	**Clean EDA Change**	
H 2a	0.0030	0.8225	0.8083	ANOVA
	0.0046	0.8228	0.8086	t-Test
H 3a	0.5496	0.4913	0.4317	ANOVA
	0.5496	0.4915	0.4320	t-Test
H 4a	0.4356	0.6217	0.6041	ANOVA
	0.4356	0.6221	0.6047	t-Test
H 5a	0.5446	0.5330	0.5056	ANOVA
	0.5446	0.5330	0.5056	t-Test

Table 5: Summary of p-values for each hypothesis

Observing the p-values for ANOVA and t-Tests on EDA Change and Clean EDA Change, it is evident that for hypothesis 2a both 0.823 and 0.808 are drastically above the cut-off p-value of 0.05 which means that we fail to reject the null hypothesis, since there is a weak evidence against it. At the same time, the p-values for mean EDA are 0.003 and 0.004, which is low enough to indicate a strong evidence against the null, allowing us to reject it. Interestingly this is the only instance in this experiment where the null is rejected.

Looking at the p-values for hypothesis 3a, it can be observed that all three p-values of 0.55, 0.49 and 0.43 are significantly above 0.05. This means that we fail to reject the null hypothesis as there is weak evidence against it.

Moving to hypothesis 4a, the p-values for EDA Change and Clean EDA Change are 0.62 and 0.61 respectively, which again shows weak evidence against the null hypothesis meaning we fail to reject it. The result for this hypothesis stands out among hypotheses 2 to 5 as it is the only one showing a decrease in EDA; it therefore should be further discussed in subsequent sections using existing findings and theories.

Finally, looking at the p-values for hypothesis 5a, we observe across all three measures to be just above 0.5 (mean EDA: 0.55, EDA Change: 0.53, and Clean EDA Change: 0.51). Once again, this is way above the cut-off of 0.05, meaning that there is a weak evidence against the null hypothesis and that we fail to reject it.

Although the changes in the results are too small to be statistically significant, the researchers believe they are nonetheless valuable and interesting contributions and implications can be derived from them. These will be covered in the Discussion.

4.3.2 Regression

In statistics, regression analysis aims to provide the relationship between a dependent variable and one or more independent variables or covariables. Regression can only provide credible information and relation in a fixed dataset. To use regressions for prediction or to infer causal relationships, respectively, a researcher must carefully justify why existing relationships have predictive power for a new context or why a relationship between two variables has a causal interpretation. The latter is especially important when a researcher hopes to estimate causal relationships using observational data. (Freedman 2009)

Excel analysis tool pack add-in was once again proven useful as it has a regression function as well. It simplifies the process as only a few outputs must be given for the complete analysis, and only the interpretation is left for the user. The linear regression equation is:

$$y = bx + a + \varepsilon$$

Where y is the dependent variable, which we are interested in.

* *x* is the independent variable or variables, as there can be more independent variable.
* *b* is the slope of the regression line, it determines which is the rate that x changes y.
* *a* is the Y intercept, it is the point on a regression graph where the line crosses the Y axis.
* *ε* is the random error term, which is the difference between the actual value of a dependent variable and its predicted value. (Cheusheva, 2018)

The output of this function in Excel looks like on Figure 18 below. The "Regression Statistics" table in the figure below explains how well the linear regression equation fits the input dataset. The absolute value of multiple R explains the strength of the relationship, which can be positive and negative as well, having positive or negative impact on the dependent variable respectively. R squared and adjusted R square are responsible for showing how well does the data fit on the line. A higher value means more points are lose to the line. R square of 0,95 or over is considered a good fit. Adjusted R square is adjusted by the number of independent variables. Standard error is an absolute measure that shows the average distance of data points from the line, which means a smaller number means a better fit. The ANOVA part was already discussed in the previous chapter. It is rarely used in linear regression analysis. The last table in the figure below shows the coefficients and provides information about the specific components of the equation. It enables linear regression equation in Excel, as seen on Figure 19 below on the next page.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.042896242
R Square	0.001840088
Adjusted R Square	-0.03143191
Standard Error	3.593242912
Observations	32

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.714056829	0.714057	0.055304	0.815675565
Residual	30	387.3418388	12.91139		
Total	31	388.0558957			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.002015961	1.433184084	2.094648	0.04475	0.075063581	5.928968342	0.075063581	5.928968342
Duration Seconds	-0.000167675	0.000712997	-0.23517	0.815676	-0.001623809	0.00128846	-0.001623809	0.00128846

Figure 18: Regression output in Excel

As the two factors identified in the methods for evaluating a decision is the timing of the decision (measurement of the time) and decision quality (measured by the scoring system) regression provides a great opportunity to consider these two factors at the same time. The unit of time was changed from minutes to seconds here as well to provide a more precise reading. The first regression was running the tests duration against the mean EDA as a dependent variable, which is visible on figure y above.

As mentioned earlier in an optimal case, Adjusted R square should be above 0,95, which would mean 95%+ percent of the data fits the trendline, which is not the case. The coefficient of each independent variable gives the slope of the line, which currently is -0,0002 for the Duration. The coefficient of the intercept is 3,002, which means the final equation is y=-0,0002x+3,002. Unfortunately, the P-value, which tests the null hypothesis that the coefficient is equal to zero (no effect) is high for the duration, which means the null hypothesis cannot be rejected. The equation enables to visualize the data on a graph with the trendline included as seen below on Figure 19.

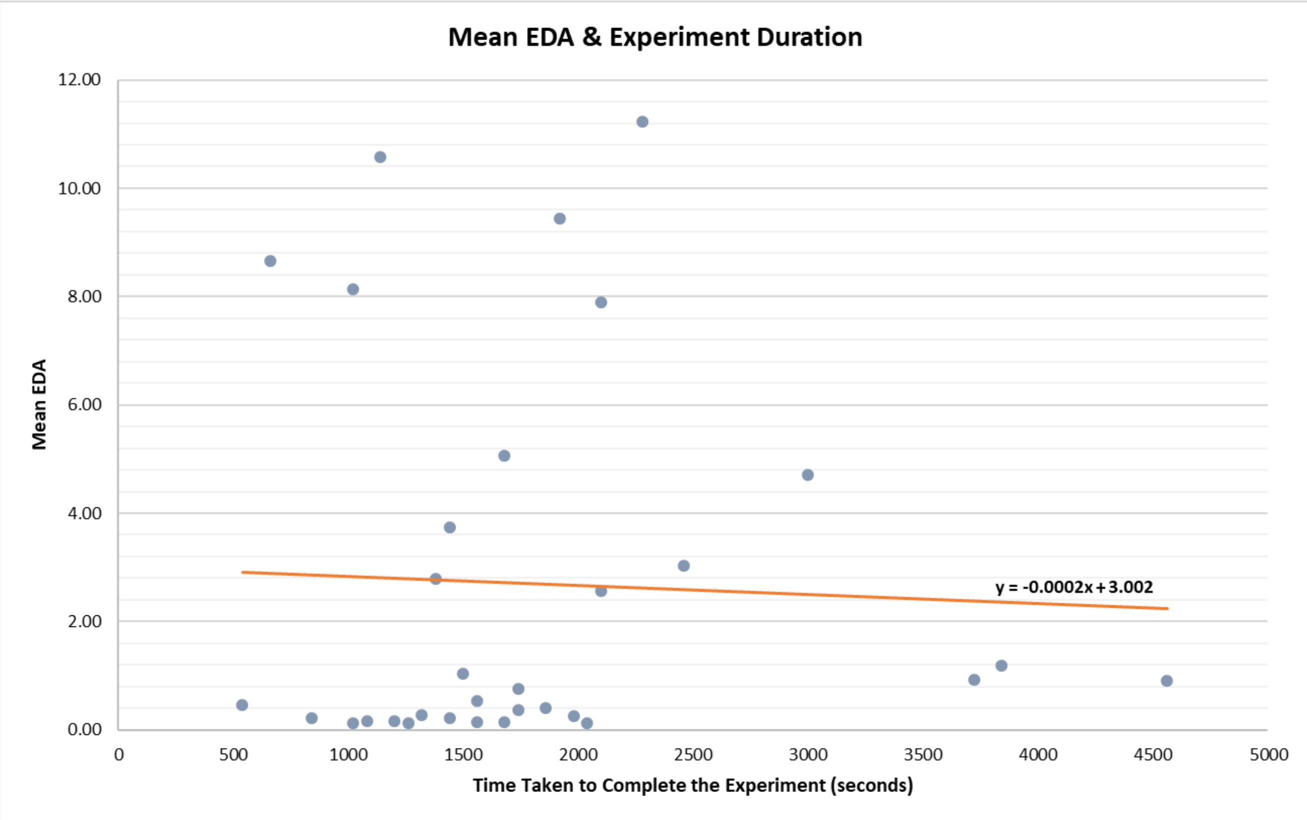


Figure 19: Linear Regression scatter plot of Mean EDA and Experiment Duration (in seconds)

In the next step the hypotheses are included in this regression separately on top of the duration according to the table below in Figure 20. Hypotheses are included in separate columns and are marked in the column using values 1 or 0 based on whether that version included the hypotheses changes or not respectively.

	A	B	C	D	E	F	G	H	I
1	**Participant number**	**Excel Version**	**Average EDA**	**Perceived CL**	**Duration Seconds**	**Hypothesis 1**	**Hypothesis 2**	**Hypothesis 3**	**Hypothesis 4**
2	**1**	*0010	0.152	3.000	1680	0	0	1	0
3	**2**	*0100	4.704	3.250	3000	0	1	0	0
4	**3**	*0000	2.572	3.750	2100	0	0	0	0
5	**4**	*0001	0.376	4.500	1740	0	0	0	1
6	**5**	*0110	0.120	3.500	1260	0	1	1	0
7	**6**	*1000	10.573	3.500	1140	1	0	0	0
8	**7**	*0101	2.798	3.750	1380	0	1	0	1
9	**8**	*1011	11.226	3.333	2280	1	0	1	1
10	**9**	*1001	0.271	3.750	1320	1	0	0	1
11	**10**	*1010	0.414	3.000	1860	1	0	1	0

Figure 20: Data Used to run Regression Tests

Table 6 below shows the regressions run by including duration and each hypothesis as an independent variable to Mean EDA as a dependent variable. As previously discussed, according to the other tests the first hypothesis looked valid. It also has a higher than average Adjusted R square value, but still not close to the expected 0.95 or above. The coefficients of duration are all similar to the values of the previous regression, while the the hypothesis values each are very different. H1 has a steepness of almost 4, h2 and h4 are similar at around 0,76 and h3 has the lowest at -0,1,12. The P value for hypothesis 1 is once again below 5% but even the second lowest value is above 40% for hypothesis 3.

Dependent Variable	Independent Variables	Adjusted R-Squared	Coefficients	P-Value
Mean EDA	Duration Seconds	0,259898101	0,000178	0,7728
	Hypothesis 2a	-	3,908111	0,0012
Mean EDA	Duration Seconds	-0,054292259	-0,000160	0,8264
	Hypothesis 3a	-	0,759464	0,559
Mean EDA	Duration Seconds	-0,040934398	-0,000306	0,68
	Hypothesis 4a	-	-1,115101	0,4011
Mean EDA	Duration Seconds	-0,053760733	-0,000169	0,8167
	Hypothesis 5a	-	0,775049	0,5508

Table 6: Regression values for Mean EDA, Duration Seconds and Hypotheses 2a to 5a

The same regression test was running on EDA Change (relative comparison of last 100 and first 100 valid values) and Clean EDA Change (EDA Change with just the two middle quartiles included) as a dependent variable. The adjusted R square variables still show little connection with the trend line. The coefficients are a lot closer in value and the P-values are still extremely high, which indicates weak to no evidence. The numbers are consistently inconsistent for all hypotheses with the new indicators. As it is the case, the original mean EDA was proven to be a better indicator for the research.

Dependent Variable	Independent Variables	Adjusted R-Squared	Coefficients	P-Value
EDA Change	Duration Seconds	-0,061440542	0,000137509	0,6597
	Hypothesis 2a	-	0,086303467	0,8768
EDA Change	Duration Seconds	-0,044898251	0,00013386	0,6617
	Hypothesis 3a	-	0,375362488	0,4922
EDA Change	Duration Seconds	-0,056306474	0,000101805	0,7466
	Hypothesis 4a	-	-0,226279479	0,6871
EDA Change	Duration Seconds	-0,048385723	0,000129473	0,6726
	Hypothesis 5a	-	0,335676248	0,5393

Dependent Variable	Independent Variables	Adjusted R-Squared	Coefficients	P-Value
Clean EDA Change	Duration Seconds	-0,061240538	0,000136431	0,6646
	Hypothesis 2a	-	0,114697269	0,838
Clean EDA Change	Duration Seconds	-0,040182713	0,000130859	0,6706
	Hypothesis 3a	-	0,430897359	0,4336
Clean EDA Change	Duration Seconds	-0,055880357	0,000095978	0,7625
	Hypothesis 4a	-	-0,244304031	0,6662
Clean EDA Change	Duration Seconds	-0,04690003	0,000125851	0,6834
	Hypothesis 5a	-	0,361213336	0,5122

Table 7: Regression values for EDA Change, Clean EDA Change and Hypotheses 2a to 5a

4.3.3 Perceived Cognitive Load

The EDA measurements were tested in many ways, including t-Test, ANOVA, Regression, by being compared to the duration and each hypothesis in several methods. Other indicators were also considered, but as the expected outcome is similar, after short testing the ideas were discarded. A different approach to the data would be to change the examined data this time, instead of the testing method. As previously mentioned, the experiment generated several different outputs, some looking uninteresting in terms of the hypotheses, but an interesting approach could be to look at the survey questions at the end, more specifically the ones examining the perceived cognitive load.

In the first set of calculations the EDA Change was replaced by the average perceived cognitive load, which was generated by taking the average of the survey question answers, which were on a scale of 1-5. First this perceived cognitive load was ran as a dependent variable compared to Duration as an independent. This process was repeated with each of the hypotheses. The adjusted R squared indicator indicates that almost no amount of data is positioned on the trend line. The P-values are extremely high, with the lowest being rounded to 0,48 which means none of the null hypothesis can be discarded.

Dependent Variable	Independent Variables	Adjusted R-Squared	Coefficients	P-Value
Perceived CL	Duration Seconds	-0,023376878	-0,00005912	0,5930
Perceived CL	Duration Seconds	-0,050932653	0,00003196	0,7780
	Hypothesis 2	-	-0,11876258	0,5581
Perceived CL	Duration Seconds	-0,044908313	0,00004398	0,6937
	Hypothesis 3	-	0,14210928	0,4766
Perceived CL	Duration Seconds	-0,062762551	0,00003854	0,7385
	Hypothesis 4	-	-0,03170143	0,8774
Perceived CL	Duration Seconds	-0,062405175	0,00004251	0,7057
	Hypothesis 5	-	-0,03661775	0,8550

Table 8: Regression for Perceived cognitive load

4.3.4 Perceived Cognitive Load and Clean EDA Change

During the final regression testing, the Clean EDA Change was compared directly to the Perceived Cognitive Load in order to explain the connection between them based on the available data. Regression was used in order to explain the relation between the two indicators with Clean EDA Change being the dependent variable and Perceived Cognitive Load being the independent variable. Figure 21 below presents the Adjusted R Square which shows the relation to the trendline, the coefficient that is used in the equation for the independent variable and the P-value that is significantly above 0,05 cut-off value. Using the coefficient and the Intercept, the equation comes out as seen in Figure 21, y = 0,3067-0,615.

Dependent Variable	Independent Variables	Adjusted R-Squared	Coefficient	P-Value
Mean EDA	Perceived CL	-0.02057843	0.30667578	0.5449

Table 9: regression for Mean EDA and perceived cognitive load

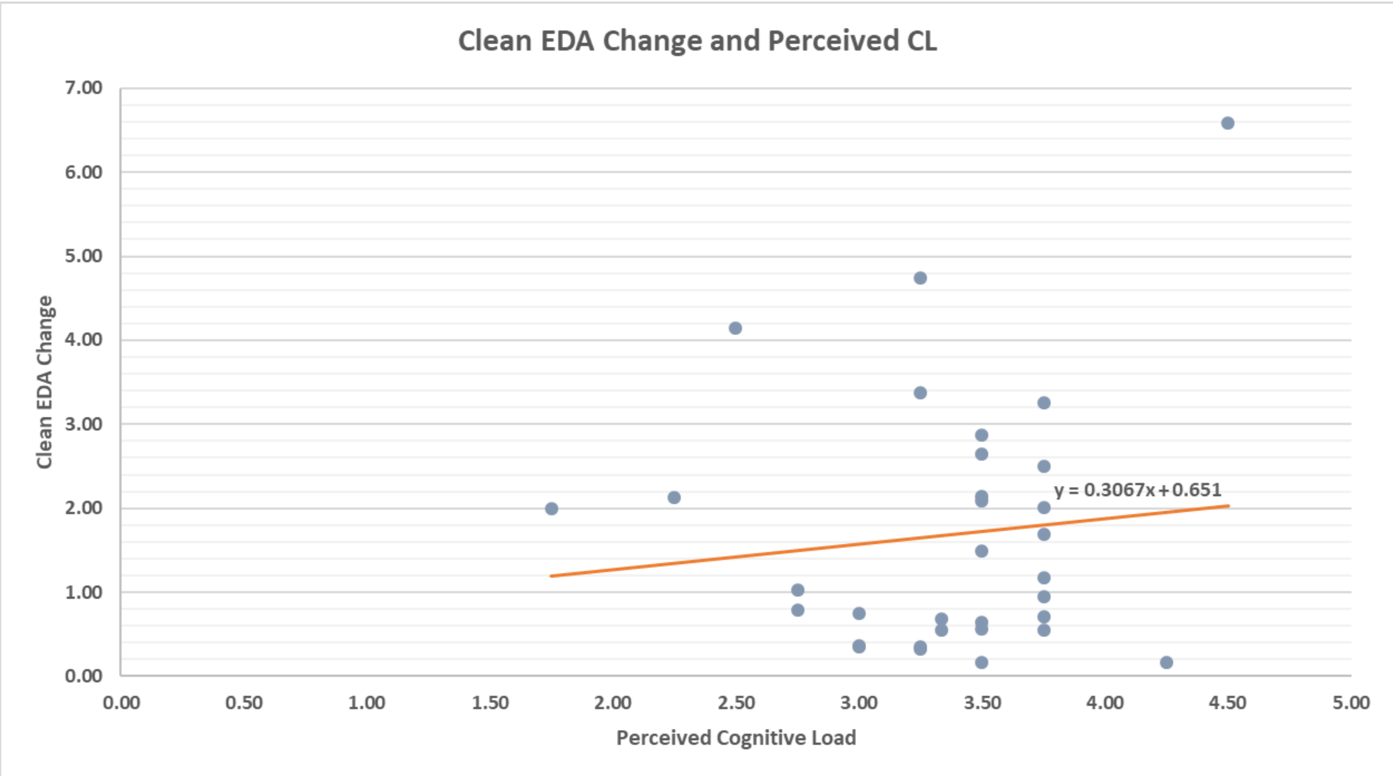


Figure 21: Linear regression scatter plot with a trendline for Clean EDA Change and Perceived Cognitive Load

5. Discussion

The next chapter introduces the discussion of the research, where the findings of the experiment are examined on a deeper level through the reflection of the literature review. First, the Findings are summarized, as it can get a bit too technical in the last chapter, then the discussion is following, first summarizing the major contributions of the research, then based on the hypotheses, describing if they were supported or not from different perspectives and providing possible explanation why, or why not in each case. Based on this, implications are defined for both theoretical and practical purposes, then the limitation of the current paper and possibilities of future research in this area, based on this research are discussed.

5.1 Summary of Findings

The findings section started with the description of the measurement process, the obstacles encountered throughout the process and the final outputs of the experiment. These outputs generated a Microsoft Excel file where different variables or indicators were determined, which were to be used later for the analysis. The first of these indicators was the „mean EDA" which was the average of the readings of the wristband throughout the whole process. It was complemented with the corresponding standard deviation values for each individual. It was considered a decent starting point for analysis, but not perfect, as there were several outliers in the dataset. In order to provide a better analysis, the „EDA change" of the dataset was provided throughout the process. In order to determine this, the average of the second 100 and the last hundred values of the measurement was calculated, and distributed with each other. This generated a proportional value with reasonable maximum range in the dataset. In order to further clean this data up, further calculations were made, and the lower and upper quartiles of these values were excluded from the averages, generating the „clean EDA change" indicator, which was considered the main datapoint of the EDA measurement. Another important aspect was the timing of the decision. The „duration" of the test also became an important indicator for further analysis, first measured in minutes to provide a consumable number for descriptive analysis, and then later changed the format to seconds to have more accurate calculations. „Decision quality" was determined by the expert's opinion through the researches own pointing framework, which is an important indicator, and both in the reflection of time, which was named „Points per minute". The last of these variables is the value of „perceived cognitive load", which was determined through the average of self-reflected cognitive load questions at the end of the survey, having a range of answers from 1-5. These variables provided the opportunity to advance further and start the analysis.

The analysis started with a descriptive process grouped by the hypothesis. First, the data groups of mean EDAs were presented grouped by the hypothesis. Each hypothesis was shortly described, and the correlated indicators were presented for them. A graph was made for hypothesis 2 to 5, which presented the „mean EDA", „EDA change" and „clean EDA change" of the groups, distributed by whether they contained the hypothesis or not, complemented by their respective standard deviation, using box plots, made in Microsoft Excel. As the second part of the hypothesis, their direct impact on decision quality was also described with their correlated data. It was repeated for all of hypothesis 2,3,4 and 5. The first hypothesis deserves its own section, as it is a global hypothesis that encompasses the whole research and the other hypotheses. It was first described with a graph, which presented the „clean EDA change" and the points per minute, for each individual, which shows that people with highest points per minute values had low amount of „clean EDA change". A scatter plot with a trendline was also provided for hypothesis 1 to visualize the data. Next the mean of points was grouped by the hypotheses and presented in a previously mentioned box plot. H1 and h3 resulted in a higher decision quality, while h2 and h4 resulted in lower.

As the chosen research philosophy was post positivism (or critical realism), it states that there is one reality, but it is hard to grasp and comprehend, and it requires to examine from many different angles. Post hoc analysis, focuses on alternative ways to analyse the data if the initial research plan doesn't prove to be enough. In this paper, post hoc analysis contains the statistical calculations. Initially ANOVA and t-Test were ran on "clean EDA change" to show statistically significant results for hypothesis 2-5. As it was not the case, further methods had to be attempted. Regression provided a way to test based on several independent variables, and to provide trendlines for these variables on scatter graphs. All of mean EDA, EDA Change, Clean EDA Change and perceived cognitive load were ran against duration in seconds in itself, and each of the hypothesis separately combined with duration, but unfortunately it didn't result in statistically significant data either. Another promising perspective was to compare self-reported cognitive load to the EDA measurements, so regression was run for mean EDA, EDA change and clean EDA change as well, against perceived cognitive load and trend lines were created to represent these relationships.

This summarizes the findings of the research. As expected, statistically significant data was not found, but it was compensated by more in-depth analysis for the description, which was possible thanks to the sample size.

5.2 Main Contributions

In the following sections the main findings of this research are interpreted, possible explanations are proposed, and the main contributions are discussed. First the main contributions of the paper are discussed overall, reflecting on the purpose of the research, then they are done in the order of the hypotheses starting with the main hypothesis 1, followed by each pair (a and b) of hypotheses 2 to 5.

The reoccurring question throughout the research was if expert decision making can be replicated for novices through presentation of data. It is researched in order to help companies handle novices in a better way, indirectly helping them with a more pleasant and efficient learning experience. As explained before, the technological advancement enables a more widespread audience to use Business Intelligence and Analytics in their everyday life and work, but this widening is obviously including less experienced people. This, according to Kahneman and Sweller, reduces the average decision quality through increased cognitive load. The first major contribution stems from this. Companies should aim to help and support novice decision makers to enhance their learning experience, and to make scheme creation, which they can store in long term memory, easier. As stated, domain specific knowledge is the primary factor in this difference between novices and experts. The second major contribution is that the way data is presented is also a highly impactful factor, which can compensate in a proper way for the lack of experience. The third main contribution can also be connected to the previous theories of Kahneman and Sweller. It states that cognitive load is a mediating variable for decision quality. Lower cognitive load generates better decision quality and a more efficient learning opportunity in instructional design.

5.2.1 Hypothesis 1

The central hypothesis of this research is based on the assumption that decision making is generally of better quality and less stressful for experts than it is for novices. The prediction was therefore made that if a novice was in a data-based decision-making situation their intrinsic cognitive load would increase, decreasing their decision quality as a result. From the findings, disregarding some of the outliers, it can be observed that there is a slight downward sloping trend between the change in cognitive load and the points scored per minute. This means that when the participants' EDA level changes were higher their decision quality was negatively affected, thus supporting Hypothesis 1 as well as some of the theory this hypothesis was based on.

The relationship observed between the change in EDA and the points scored per minute can be attributed to several theories outlined in the Literature Review section. First, the level of expertise needs to be addressed. High changes in cognitive load could be explained by the fact that when an individual processes new information in a new way their working memory capacity can become overloaded and engaging the long-term memory, and consequently increasing the stress levels over the course of the experiment. Moreover, the participants lacked domain specific knowledge in the task they had to do, making the case questions and the dataset highly complex for them. In fact, upon completion of the experiment most of the participants said they felt overwhelmed due to the combination of complex questions and dataset. Several participants even mentioned they simply did not know what they were doing or how to answer some of the questions. This supports the theory that novices do not have cognitive structures called schemas developed in their long-term memory, meaning they have none or little knowledge as to how to solve the problem or what information to look for.

Although the hypotheses are supported, if we look at individual cases, some interesting observations can be made. For example, certain individuals scored high both in terms of total points and points per minute despite their high change in EDA. Similarly, other individuals scored poorly despite their low EDA changes. While there is a possibility that the such high or low scores were obtained simply due to chance, the researchers came up with several potential explanations for this that haven't been foreseen. This could be demonstrating that the individuals were not necessarily experiencing a cognitive overload but rather were simply focused and actively engaged in the decision-making process. Alternatively, this could be showing the level of incentive the participants had to make good decisions; meaning that if they did not care or give it a lot of thought they would not stress or engage in the task. Both of these motives are possible and will be discussed further in the sections below. They are not mutually exclusive and raise interesting questions for future research.

The ultimate goal of this study is not only to help non experts make decisions of similar quality as experts, but also to be able to understand and comprehend the data as an expert would. A prediction was made that an effective way to achieve this, is by presenting the data in a specific way that is easy for novices to analyse. Therefore, the next four sections will look more specifically at how the four selected presentation ways can help to reduce cognitive load and increase decision making quality.

5.2.2 Hypothesis 2a and 2b

The prediction in hypothesis 2a was that including Raw Data and exposing novices to it would increase the individual's cognitive load. Similarly, it was predicted that as a consequence of an increased cognitive load decision quality would decrease. From the findings it can be observed that while cognitive load did increase as was predicted, the decision quality increased as well. Firstly, the aim with this hypothesis was to target the extraneous cognitive load by adding an element that distracts the participant and creates the "split-attention effect" mentioned in the literature review. Presence of pure Raw Data without any additional graphs was theoretically supposed to expose the participants to more stimuli and make the task more complex. Recalling the literature, exposing an individual to more stimuli can cause an overload of their working memory, while complex tasks can cause the participants to make a decision that is not necessarily correct but is satisfactory enough. This could cause them to spend more working memory resources on processing information and therefore increasing their stress level. Although higher changes in cognitive load can be explained by these theories, findings for hypothesis 2b contradict this conclusion.

Higher decision quality in combination with a higher change in EDA could be suggesting that this level of cognitive load is not exceeding the limit and does not lower decision quality. Once again, as mentioned above, this could be explained by the fact that the individuals were more concentrated on the task rather than stressed and overloaded by it. In fact, some additional interesting research has been found to support this claim. Sörqvist et. al. (2016) argue that in a task that requires visual attention the neural response to other stimuli, in particular auditory stimuli, is suppressed "along with increased activity in networks related to effortful attention". This suggests that higher cognitive load could indeed mean higher focus on a given task and a decreased level of distractibility.

What this means in the workplace is that when experts present data to novices, including raw data will not harm their decision quality. Furthermore, it could even increase the concentration and engagement of novice employees in the tasks given to them as well as ultimately speed up their learning processes and growth at the company. This effect in the workplace, however, should be further investigated. Another important consideration to make is the perceived cognitive load, as many participants stated they felt overwhelmed by the amount of data and complexity in the task. Management and experts need to keep this in mind when preparing data for novice employees, because even if their actual cognitive load is not excessively high, they may feel too stressed which could consequently affect their overall job satisfaction.

5.2.3 Hypotheses 3a and 3b

Hypothesis 3a predicted that instructing participants to engage in extra decision making in terms of data visualisation would increase their cognitive load, more precisely their extraneous cognitive load. Additionally, in hypothesis 3b it was predicted that an increase in cognitive load would consequently cause a decrease in decision quality. The findings show that the predictions were correct and both hypotheses are supported. In this case the findings support and can be explained by the theories discussed in the literature.

Higher increase in cognitive load could be attributed to the “the redundancy effect” described by Sweller (2002). As previously mentioned, he argues that redundant activities, in the case of this hypothesis adding extra visualisations or even simply using the computer or unknown software, could cause an overload of the working memory. In terms of decision quality, it was stated that the quality of the information on which the decision was based influences quality of that decision. As was assumed, novice users do not have the necessary domain knowledge, and they have not developed the necessary cognitive schemas and pattern recognition skills. Due to this lack of knowledge, skills and understanding, asking them to manipulate data could impede the quality of that data and consequently the quality of the decisions made based on that data.

On a company level this offers an important contribution for managers and experts. That is, in a situation where a novice employee needs to make decisions based on a dataset prepared by the expert, the expert should provide a final dataset with the necessary graphs and calculations rather than expect the novice to create those on his or her own. Having more experience, the expert should know what information and data is required to make specific types of decisions while a novice may not. Immediately providing the novice with all the necessary data could reduce their cognitive load, improve their decision making and even improve their learning process.

5.2.4 Hypotheses 4a and 4b

The prediction in Hypothesis 4a was that presenting data using multiple data visualisation formats will increase cognitive load, more precisely the extraneous cognitive load. At the same time, hypothesis 4b predicted that this increase in cognitive load would consequently decrease decision quality. The findings of this experiment show that the participants actually experienced a decrease in cognitive load and an increase in decision quality. It can be argued that being exposed to different visualisation formats decreases stress levels in individuals and helps them to better understand the data and ultimately make better quality decisions. Although it was originally predicted that presenting multiple visualisation formats would constitute as multiple stimuli and overload the participants’ working memory, and consequently decrease their decision quality. The findings however contradict this theory.

A possible explanation for this could be that visualising the data in different ways actually helped novices to differentiate between the various elements of the dataset and more easily navigate through it. This would also result in a better understanding and an increased decision quality. The implications this provides for the workplace is that managers or experts should consider which format fits which type of data best to result in better understanding and distinction. They should subsequently provide novices with the datasets containing graphs that are appropriate for the type of data being used and analysed.

5.2.5 Hypotheses 5a and 5b

Hypothesis 5a predicted that colour coding the data using multiple colours will increase cognitive load, which consequently leads to the prediction in hypothesis 5b that increased cognitive load decreases decision quality. Similarly, to hypothesis 4a the prediction was based on the theory that exposing an individual to too many stimuli will cause a cognitive overload. Therefore, the standard conditional formatting function in Excel was used to colour code the data, ensuring that over seven colours was used for the different elements.

From the findings it can be observed that as predicted participant cognitive load increased at the same time as decision quality decreased. This could be attributed to the theory discussed in the literature review. For example, it was discovered that the working memory has a limited capacity, in particular when processing unfamiliar information an individual can focus on at most 7 elements at the same time. In this case, colour coding the dataset with over 7 colours resulted in a cognitive overload in the participants. The participants simply did not understand what the colours meant, and it was too overwhelming to figure out what they meant. Not being able to easily compare the various elements of the case could have been a determining factor in reducing their decision quality.

These two hypotheses provide the following managerial implications; first of all, managers or experts should not apply an excessive number of colours when preparing data for novices. Although some colour coding could be beneficial, it would be recommended to use less than 7 colours. However, the precise number should be investigated further.

5.3 Implications for practice

Based on this research, several important implications for practice can be derived. First of all, professionals need to realise and acknowledge that there are important differences between experts and non-experts when it comes to data-driven decision making. These differences can be observed in the lack of domain specific knowledge and pattern recognition, as well as increased levels of stress and cognitive load. This can consequently result in poor decision quality.

Management should recognise the need to adapt certain information or datasets for the novice employees in a way that is appropriate and relevant for their level of data processing and analysing skills. This will help to not only reduce stress and cognitive load but more importantly to enhance their understanding of that information. Reduced intrinsic and extraneous load will highly likely increase the germane load and benefit the employees' learning curve as well as their overall wellbeing, inevitably benefiting the company long term.

This research looked at four specific types of data presentation that could affect the cognitive load and decision quality of non-experts. These were: presence of raw data, additional visualisation decisions, excessive visualisation formats and excessive colours. It was discovered that excessive colours and the need to make additional visualisation decisions increase cognitive load and decrease decision quality. This means that when preparing data for novices limited number of colours should be used and it needs to be clear what the colours mean. Additionally, all the data and graphs that the novices need should be included in the dataset. This will prevent them from engaging in additional actions and will ensure a smooth learning process, so that these additional data visualisation decisions can be slowly introduced as the individual learns and obtains the ability to recognise patterns as experts do.

Simultaneously, it was discovered that including raw data in the dataset and presenting it using multiple visualisation formats will not necessarily result in cognitive overload or decreased decision quality. With regard to presenting raw data, this could in fact increase the germane cognitive load resulting in better and faster learning and therefore better decision quality. With regard to multiple visualisation formats, on the other hand, this emphasises the fact that data needs to be presented in an appropriate and relevant way. Visualising data in the right way could help novices to read it better and navigate or distinguish more easily between the various values.

5.4 Limitations

The research was conducted in the boundaries of Copenhagen Business School, which somewhat also determines the timeframe available for the research and the location of easy access information sources in Europe, and even more specifically, Denmark. It is important to mention, that this is not considered a limitation in a negative sense, it is more like the description of the context of this paper. Time was neither considered a factor that limited the preferred research methods in any way.

The focus originates from BI decision making and human behaviour, which suggested the distinction between novices and experts based on the available theories. As researching this area of human decision making is continuously changing according to the advancing technology, and is still in an early stage overall, a more generalized overview of the area was chosen. However, a perspective had to be in focus, and the company's or expert's perspective was chosen, because it can indirectly benefit the novices as well, which pivoted the question towards: How can companies design their dataset in order to improve decision quality by novices, instead of how novices can aim to make decisions just like experts.

The research was somewhat limited by experience, as the researchers had limited previous experience in conducting quantitative experiments, using scientific measurement methods and equipment and using statistical calculations. The context of master thesis plays an important part here, as this opportunity was viewed as a great learning opportunity, and the boundaries of the thesis provided the necessary help in order to conduct the research with great success.

The necessary equipment was provided by Copenhagen Business School. The researchers have never used or seen the technology before, but it could be easily compared to a smartwatch that is getting more and more widespread these days. The accuracy of the equipment, as previously mentioned, was worse than expected, which had to be taken into consideration in various ways later into the research.

The experiment was conducted with 33 participants, which sounds like a low number for statistical analysis, but the decision was made, based on several factors, to go with this number. Relevant literature and other highly credible researches were conducted with the same amount of people with similar methods and the number of hypotheses made it suitable as well. It provided the opportunity to go for a more in-depth analysis for the test. On the other hand, the time limitation played a crucial part, as just the actual time it took for the test was more than 16 hours over 2 weeks, excluding preparations for the test or organization of appointments or finding participants, which was the bigger part of the process. The hypotheses of this research were formed in a way to provide opportunity for further research. A more defined scope for each of the presented data characteristics can provide opportunity to research them on a deeper level.

Finding participants for the experiment was proven to be easier than expected, but still difficult. As mentioned before, an opportunistic approach had to be taken, but it didn't limit the research, as the boundaries for the sample were broad, just as the focus, to provide an overall picture of decision making of companies.

5.5 Future research

As previously described in the last chapter, the research topic holds great potential. There haven't been many previous studies done in this area, and continuous technological advancement changes it rapidly over the years, the potential to expand this knowledge and narrow the scope is present. This paper can provide a research basis for further studies related to the relation of decision making and cognitive load.

One way to elevate this research, is lifting the limitation of this paper. The time limitation is not viewed as an obstacle, but the location strongly influences the cultural background of the study. Different countries, or even more preferably other continents could show interesting differences based on cultural differences. Another difference could be the change of perspective, focusing on the novice side, how can they collect experience in a more efficient way to transition into being an expert. This topic of efficient learning was researched by many researches through the Cognitive Load Theory. Another change could be the change of sample size. A bigger sample size could possible provide statistically significant differences, but as previously mentioned, EDA measurement if a complicated process with several components, makes it less reliable than desired, but with caution, feasible.

The other way to expand the research is changing the focus to more narrow area of this research. As this is a complex process, many components can be slightly more specialized in order to show a deeper picture. The most obvious focus could be focusing on a specific industry. In this paper the company was chosen based on the available opportunity, and the industry was not specified. These decision situations can provide a vastly different outcome for the research. The participants can be also chosen more precisely, based on the researched area, if it’s necessary, both novices and experts. The experts could be chosen more strictly based on their position at the company, and the novices could be chosen based on their specific study program or something similar defining trait.

Decisions can be categorized in many ways, which provides another way to go into the specifics. Decision centralization, how is the decision quality between centralized or decentralized decisions? In the Business Intelligence and data-driven decision-making chapter, it was mentioned that decisions are made through the whole company on all levels, and different problem structures, structured, semi-structured or unstructured decisions, are also a possibility to differentiate from the general overview. The involvement of BI technologies can also be increased. In the end, this paper uses Microsoft Excel as a BI tool, which on the one hand is not the most defining tool of BI, but on the other hand, several companies still use it for their BI functions, undoubtedly.

The technological evolution plays an important part in this process. Cognitive decision support systems have been mentioned several times in the literature, and with the recent surfacing of AI technologies, it opens the way for more opportunities to improve and to research it.

6. Conclusion of discussion

The human cognitive and learning abilities are the outcome of an ongoing process that started tens of thousands of years ago. Considering the vast amount of data and information the world currently holds, and the continuous advancement of mankind, humans must learn throughout all their lives. Technology is not an exception in this regard, information society is advancing at an incredible rate, and the efficient use of data and information is required for success. Through various researches and theories, Cognitive Load Theory by John Sweller (1988) concluded the benefits or learning, by stating that domain specific knowledge is the primary factor distinguishing novices from experts in terms of problem-solving skills. This research examined this statement in a Business Intelligence setting with the goal of providing a way for companies to present their data in an appropriate way for novices to enable them to make better quality, expert like decisions. An experiment was conducted examining four types of data presentation in a data-driven decision-making setting, capturing an EDA measurement of the participants. This formed the basis for the research and resulted in interesting findings and valuable implications for practice. The main hypothesis was formed in a deductive way, testing the statement of the studied literature that higher cognitive load results in a lower decision quality. Four other hypotheses were formed by looking at the characteristics of data presentation, namely raw data inclusion, visualization decisions, visualization format and colour coding were tested in terms of cognitive load. Based on the outcome of the experiment, it can be concluded that the first hypothesis is support, as in a higher cognitive load results on a lower decision quality on average. Reflecting back on the research question, certain aspects of data presentation can be used in a beneficial way to lower the cognitive load and indirectly increase the decision quality, as shown by the examples in each of the hypotheses.

Bibliography

Al-Tarawneh, H. A. (2012). The main factors beyond decision making. Journal of Management Research, 4(1), 1-23.

Anthony, R. N. (1965). *Planning and control systems: A framework for analysis [by]*. Division of Research, Graduate School of Business Administration, Harvard University.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Bhushan, N., & Rai, K. (2004). Applying the Analytical Hierarchy Process. Strategic Decision Making.

Boss J. 2015 The 9 Characteristics of a Good Decision [Accessed 2019 September] https://www.entrepreneur.com/article/250175

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology*, *49*(1), 1017-1034.

BusinessDictionary. (2019). What is decision making? Definition and Meaning. Retrieved 2019, [Accessed September-October 2019] from http://www.businessdictionary.com/definition/decision-making.html.

Conrad A, 2019 BI vs Big Data vs Data Mining: A Comprehensive Comparison of the Difference Between Them [Accessed 2019 September] https://selecthub.com/business-intelligence/bi-vs-big-data-vs-data-mining/

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS quarterly*, *36*(4).

Cheusheva S. 2018 Linear regression analysis in Excel [Accessed 2019 September] https://www.ablebits.com/office-addins-blog/2018/08/01/linear-regression-analysis-excel/

DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in human behavior*, *23*(3), 1616-1641.

Dewey, J. (1910). How we think. Boston: Heath.

Dewey, J. (1916). Method in science teaching. Science Education, 1, 3-9.

Empatica 2019 [Accessed 2019 September]
https://www.empatica.com/connect/dashboard.php

Endsley, M.R., Bolte, B., Jones, D.G.: Designing for situation awareness: an approach to user-centered design, 1st edn. Taylor&Francis, London (2003)

Evaristo, R., Adams, C., & Curley, S. (1995). Information load revisited: A theoretical model. ICIS 1995 proceedings, 18.

Evelson, B., Moore, C., Kobielus, J., Karel, R., Nicolson, N.: The Forrester wave: Enterprise business intelligence platforms, Q3 (2008)

Freedman, D. A. (2009). *Statistical models: theory and practice*. Cambridge university press.

Gartner Inc. a), 2017. *Gartner IT Glossary - Digitalization.* [Accessed 2019 September] https://www.gartner.com/it-glossary/digitalization/ 17

Gartner Inc. b), 2017. *Gartner IT Glossary - Digitization.* [Accessed 2019 September] http://www.gartner.com/it-glossary/digitization/

Geary, D. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1–99). Greenwich, CT: Information Age Publishing.

Gruman, G.: Rethinking business intelligence. InfoWorld 29(14), 22–27 (2007)

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.

Gutcheckit, 2018 A history of data collection, storage and analysis [Accessed 2019 September] https://www.gutcheckit.com/blog/a-history-of-data/

Groot, A. D. (1966). Perception and memory versus thought: Some old ideas and recent findings. *Problem solving*, 19-50.

Hal Varian 2008, Interview answer, [Accessed 2019 September] https://www.aei.org/carpe-diem/google-chief-economist-hal-varian-on-data-analysis/

Heinze J., 2014, History of Business Intelligence [Accessed 2019 September] https://www.betterbuys.com/bi/history-of-business-intelligence/

Hillsberg A, 2017, Top 5 business intelligence (BI) solutions for your company [Accessed 2019 September] https://bigdata-madesimple.com/top-5-business-intelligence-bi-solutions-for-your-company/

Huie, C. P. (2016). Perceptions of business intelligence professionals about factors related to business intelligence input in decision making. *International Journal of Business Analytics (IJBAN)*, *3*(3), 1-24.

Jacoby, J. (1977). Information load and decision quality: Some contested issues. Journal of Marketing Research, 14(4), 569-573.

Johnson, J. (2014). Designing with the mind in mind., 2 Eds.

Juslin, P., & Montgomery, H. (Eds.). (2007). Judgment and decision making: Neo-Brunswikian and process-tracing approaches. Psychology Press.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, *93*(5), 1449-1475.

Kahneman, D., & Tversky, A. (2013). Choices, values, and frames. In Handbook of the fundamentals of financial decision making: Part I (pp. 269-278).

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1–17.

Kimball, R., Ross, M., Thornthwaite, W., Mundy, J. and Becker, B., 2007. The data warehouse lifecycle toolkit. John Wiley & Sons.

Kirschner, P. A. (2002). Cognitive load theory: implications of cognitive load theory on the design of learning. Learning and Instruction, 12(1), 1–10.

KPMG 2018 Data is the new currency [Accessed 2019 September] https://home.kpmg/im/en/home/insights/2018/11/data-is-the-new-currency.html

Krettek A, 2019 The data processing evolution: A potted history [Accessed 2019 September] https://www.itproportal.com/features/the-data-processing-evolution-a-potted-history/

Kumar A, 2014 Data Presentation in Business Intelligence [Accessed 2019 September] https://www.developer.com/db/data-presentation-in-business-intelligence.html

Laursen, G.H. and Thorlund, J., 2010. Business analytics for managers: Taking business intelligence beyond reporting. John Wiley & Sons.

Luhn, H. P. (1958). A business intelligence system. *IBM Journal of research and development*, *2*(4), 314-319.

Madden, J. M. (1962).What makes work difficult? *Personnel Journal*, 41, 341–344.

Mead, D. F. (1970). *Development of an equation for evaluating job difficulty*. Brooks, TX: Air Force Human Resources Laboratory.

Medley G. 2016 Introduction to Cognitive Business Intelligence [Accessed 2019 September] https://blog.aresprism.com/introduction-to-cognitive-business-intelligence

Microsoft 2019, BI capabilities in Excel and Office 365 [Accessed 2019 September] https://support.office.com/en-us/article/bi-capabilities-in-excel-and-office-365-26c0548e-124c-4fd3-aab3-5f64568cb743

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Moray, N., Johanssen, G., Pew, R., Rasmussen, J., Sanders, A., & Wickens, C. (1979). Final report of experimental psychology group. In Mental workload (pp. 101-114). Springer, Boston, MA.

Moreno, R. E., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories.

Niculescu, A., Cao, Y., & Nijholt, A. (2009, July). Stress and cognitive load in multimodal conversational interactions. In *International Conference on Human-Computer Interaction* (pp. 891-895).

Niu, L., Lu, J., & Zhang, G. (2009). Cognition-driven decision support for business intelligence. *Models, Techniques, Systems and Applications. Studies in Computational Intelligence, Springer, Berlin*.

Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012, November). Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference* (pp. 420-423). ACM.

Patel, Salma. *http://salmapatel.co.uk.* 2015. [Accessed 2019 September] http://salmapatel.co.uk/academia/the-research-paradigm-methodology-epistemology-and-ontology-explained-in-simple-language/

Parsons J, 2016 Seven Characteristics of a Good Decision [Accessed 2019 September] https://agecon.unl.edu/cornhusker-economics/2016/seven-characteristics-good-decision

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218.

Press G., 2015 A Very Short History of Digitization [Accessed 2019 September] https://www.forbes.com/sites/gilpress/2015/12/27/a-very-short-history-of-digitization/#293d5cbc49ac

Raghunathan, S. (1999). Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. Decision support systems, 26(4), 275-286.

Resnick, M.L.: Situation awareness applications to executive dashboard design. Paper presented at the Human Factors and Ergonomics Society 47th Annual Conference (2003)

Rosedahl, J. L. (2016). Business intelligence: Strategies for improving BI adoption (Order No. 10125143). Available from ABI/INFORM Collection; ProQuest Dissertations & Theses Global.

(1807952749). [Accessed 2019 September] https://search-proquest-com.ez.statsbiblioteket.dk:12048/docview/1807952749?accountid=14468

Saaty, T. L. (1990). How to make a decision: the analytic hierarchy process. European journal of operational research, 48(1), 9-26.

Saunders, M., Lewis, P., & Thornhill, A. (2007). Research methods. *Business Students*.

Scott T, 2019, Power BI vs Tableau: A Data Analytics Duel [Accessed 2019 September] https://technologyadvice.com/blog/information-technology/power-bi-vs-tableau/

Schwartz, B., Ben-Haim, Y. A. K. O. V., & Dacso, C. (2011). What makes a good decision? Robust satisficing as a normative standard of rational decision making. Journal for the Theory of Social Behaviour, 41(2), 209-227.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2009). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine*, *14*(2), 410-417.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007, April). Galvanic skin response (GSR) as an index of cognitive load. In *CHI'07 extended abstracts on Human factors in computing systems* (pp. 2651-2656). ACM.

Singh, S.K., Watson, H.J., Watson, R.T.: EIS support for the strategic management process. Decision Support Systems 33, 71–85 (2002)

Smuts, M., Scholtz, B., & Calitz, A. (2015, September). Design guidelines for business intelligence tools for novice users. In *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists* (p. 36). ACM.

Stanovich, K. E., & West, R. F. (2000). Advancing the rationality debate. *Behavioral and brain sciences*, *23*(5), 701-717.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. Journal of Experimental Psychology: General, 119(2), 176–192.

Sweller, J., van Merrienboer, J. G., & Paas, F. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10(3), 251–296.

Sweller, J. (1999). Instructional design. In *Australian educational review*.

Sweller, J. (2002, July). Visualisation and instructional design. In Proceedings of the International Workshop on Dynamic Visualizations and Learning (Vol. 18, pp. 1501-1510).

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning andmotivation* (Vol. 43, pp. 215–266). San Diego, CA: Academic Press.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science*, 32, 9–31.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology*, 4, 434–458.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances.

Sörqvist, P., Dahlström, Ö., Karlsson, T., & Rönnberg, J. (2016). Concentration: The neural underpinnings of how cognitive load shields against distraction. Frontiers in human neuroscience, 10, 221.

Tatomir, A., McDermott, C., Bensabat, J., Class, H., Edlmann, K., Taherdangkoo, R., & Sauter, M. (2018). Conceptual model development using a generic Features, Events, and Processes (FEP) database for assessing the potential impact of hydraulic fracturing on groundwater aquifers. *Advances in Geosciences*, *45*, 185-192.

University of Massachusetts Dartmouth. (2019). PDF. Dartmouth. [Accessed September 2019] from https://www.umassd.edu/fycm/decision-making/process/.

Williams, S. and Williams, N., (2010). The profit impact of business intelligence. Elsevier.

Wu, Jonathan (2000). The Transition of Data into Wisdom. [Accessed September 2019]. https://www.information-management.com/news/the-transition-of-data-into-wisdom?fbclid=IwAR3ZrkwLncOtvcdsAC3pojqSUMG7PyI-r5ZeV8gkSd9bHf8-uzhq3nUxFsw

Yeoh, W., Koronios, A., & Gao, J. (2008). Managing the implementation of business intelligence systems: A critical success factors framework. International Journal of Enterprise Information Systems, 4(3), 79-94. [Accessed 2019 September] doi: http://dx.doi.org.ez.statsbiblioteket.dk:2048/10.4018/jeis.2008070106

Appendices

Appendix 1: Excel version differences

Hypothesis 2 control group: unhidden Raw data sheet when Excel is opened up

Channel	Date	Users	Sessions	Revenue	Transactic	Average Order Value	Conversion rate	Value per sess	Date2	Revenue excluding V	Week	Month	Year	MTD	YTD
desktop	20190101	10827	13221	253403	257	492	1,95%	19,16670449	2019.01.01	202722,4	1	1	2019	0	1
mobile	20190101	19709	25815	187483	231	405	0,89%	7,262560527	2019.01.01	149986,4	1	1	2019	0	1
tablet	20190101	9512	12272	147674	161	458	1,31%	12,03340939	2019.01.01	118139,2	1	1	2019	0	1
desktop	20190102	13126	16456	495781	483	512	2,94%	30,1276738	2019.01.02	396624,8	1	1	2019	0	1
mobile	20190102	16139	20704	229218	277	413	1,34%	11,07119397	2019.01.02	183374,4	1	1	2019	0	1
tablet	20190102	7664	9706	290139	238	608	2,46%	29,89274675	2019.01.02	232111,2	1	1	2019	0	1
desktop	20190103	9405	11181	343927	358	480	3,20%	30,75994992	2019.01.03	275141,6	1	1	2019	0	1
mobile	20190103	11968	14482	166889	222	375	1,53%	11,52389173	2019.01.03	133511,2	1	1	2019	0	1
tablet	20190103	4894	5948	251518	164	766	2,76%	42,2861466	2019.01.03	201214,4	1	1	2019	0	1
desktop	20190104	12043	14489	353794	346	510	2,39%	24,41811029	2019.01.04	283035,2	1	1	2019	0	1
mobile	20190104	14785	18230	178471	213	417	1,17%	9,789961602	2019.01.04	142776,8	1	1	2019	0	1
tablet	20190104	6397	7961	252927	171	737	2,15%	31,77075744	2019.01.04	202341,6	1	1	2019	0	1
desktop	20190105	8373	10029	262834	260	504	2,60%	26,20739854	2019.01.05	210267,2	1	1	2019	0	1
mobile	20190105	15450	19498	182067	211	430	1,08%	9,337726946	2019.01.05	145653,6	1	1	2019	0	1
tablet	20190105	6467	8026	266183	184	723	2,29%	33,16508846	2019.01.05	212946,4	1	1	2019	0	1
desktop	20190106	10263	12469	443806	377	587	3,03%	35,59275002	2019.01.06	355044,8	1	1	2019	0	1
mobile	20190106	17056	21475	248427	273	454	1,27%	11,56819558	2019.01.06	198741,6	1	1	2019	0	1
tablet	20190106	7627	9486	279330	219	636	2,31%	29,44655281	2019.01.06	223464	1	1	2019	0	1
desktop	20190107	12401	15405	442336	405	546	2,63%	28,71379422	2019.01.07	353868,8	2	1	2019	0	1
mobile	20190107	14514	18362	170140	205	413	1,12%	9,265875177	2019.01.07	136112	2	1	2019	0	1
tablet	20190107	6653	8414	218597	190	575	2,26%	25,98015213	2019.01.07	174877,6	2	1	2019	0	1
desktop	20190108	13127	16667	488388	469	520	2,81%	29,30269395	2019.01.08	390710,4	2	1	2019	0	1
mobile	20190108	16984	21761	200233	232	430	1,07%	9,20146133	2019.01.08	160186,4	2	1	2019	0	1
tablet	20190108	7777	9894	221350	197	561	1,99%	22,37214473	2019.01.08	177080	2	1	2019	0	1
desktop	20190109	10602	12974	449481	399	563	3,08%	34,64475104	2019.01.09	359584,8	2	1	2019	0	1
mobile	20190109	12309	15484	178979	213	420	1,38%	11,55896409	2019.01.09	143183,2	2	1	2019	0	1
tablet	20190109	5574	6954	200579	176	568	2,54%	28,84368709	2019.01.09	160463,2	2	1	2019	0	1
desktop	20190110	7561	8969	272132	261	521	2,91%	30,34139815	2019.01.10	217705,6	2	1	2019	0	1
mobile	20190110	9378	11330	114865	130	440	1,15%	10,13812886	2019.01.10	91892	2	1	2019	0	1
tablet	20190110	3934	4785	133836	113	589	2,37%	27,96990596	2019.01.10	107068,8	2	1	2019	0	1
desktop	20190111	8357	9896	249537	233	534	2,36%	25,21594584	2019.01.11	199629,6	2	1	2019	0	1
mobile	20190111	10328	12538	117884	130	453	1,04%	9,402137502	2019.01.11	94307,2	2	1	2019	0	1
tablet	20190111	4508	5439	136416	105	649	1,93%	25,08108108	2019.01.11	109132,8	2	1	2019	0	1
desktop	20190112	6141	7216	223310	199	559	2,76%	30,94650776	2019.01.12	178648	2	1	2019	0	1
mobile	20190112	11647	14055	122880	155	396	1,10%	8,742796158	2019.01.12	98304	2	1	2019	0	1
tablet	20190112	4844	5831	225789	139	809	2,39%	38,72217458	2019.01.12	180631,2	2	1	2019	0	1

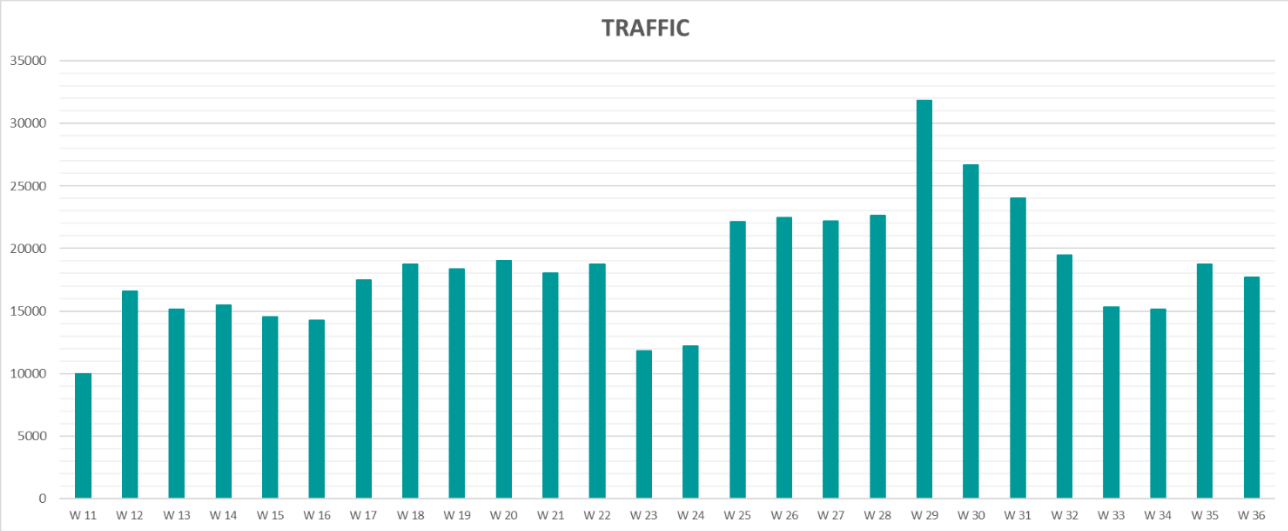
Weekly | Long term data | 2019 Averages | Responsibilities | Raw Data

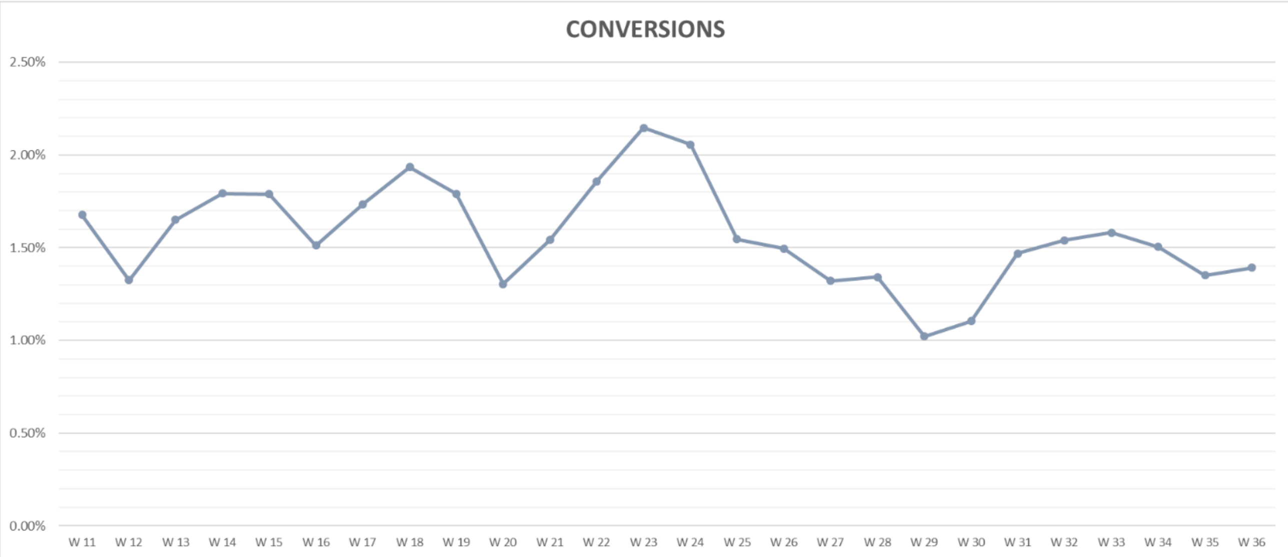
Hypothesis 3 control group: Dropdown menu is highlighted, and extra instruction was added to the case description.

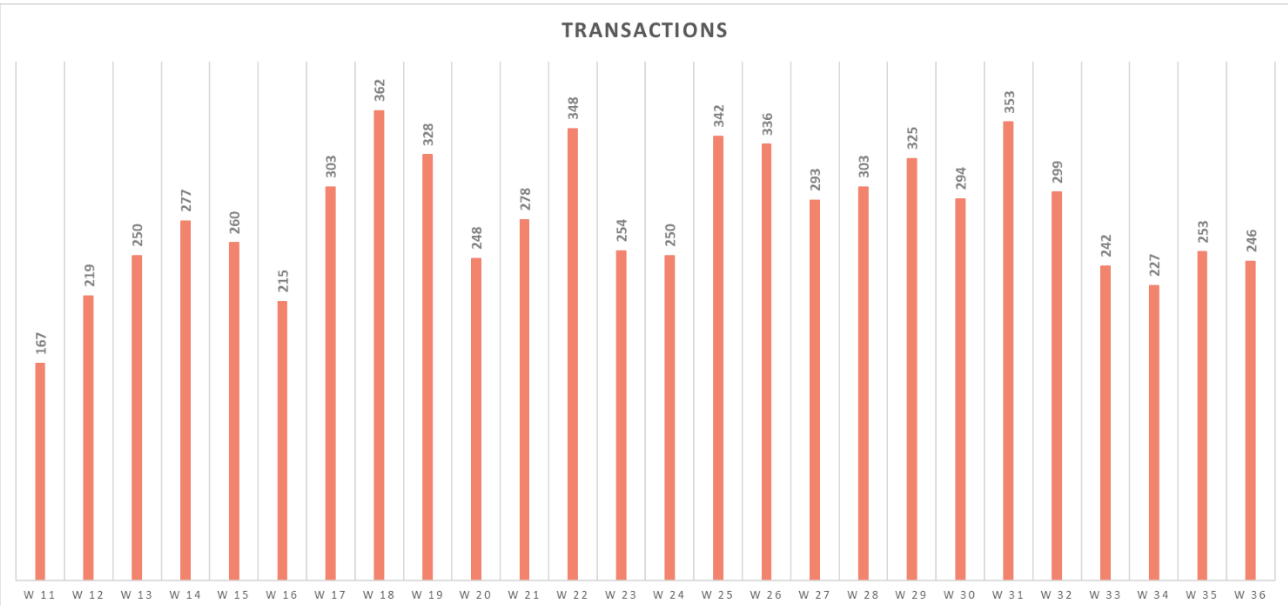
AB	AC	AD	AE	AF
	Choose the channel:			Influencers

Feel free to look through all of the sheets and add your own graphs if necessary to help you make the decisions.
Note: in sheet "Weekly" there is a drop-down list marked orange that could you help you make a decision.

Hypothesis 4 control group: Different visualization formats







Hypothesis 5 control group: Color coding function of Excel for numerical data

	Channels (Traffic)	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23
3	Direct (organic)	26,390	42,775	37,581	34,826	32,054	27,526	30,759	32,438	31,683	30,470	28,512	26,682	25,274
4	Direct (paid)	15,038	16,132	17,834	19,423	17,762	16,164	18,510	18,566	17,830	17,269	15,573	7,845	12,399
5	**Direct**	41,428	58,907	55,415	54,249	49,816	43,690	49,269	51,004	49,513	47,739	44,085	34,527	37,673
6	Google Adwords	54,037	61,640	66,435	65,874	75,598	72,860	76,561	73,268	72,923	70,946	64,267	55,697	57,761
7	Bing Ads	1,665	1,685	1,758	1,849	1,782	1,437	1,844	1,873	1,667	1,491	1,477	1,373	1,298
8	**SE PPC**	55,702	63,325	68,193	67,723	77,380	74,297	78,405	75,141	74,590	72,437	65,744	57,070	59,059
9	Google Organic	23,391	24,129	25,444	26,600	26,032	24,723	26,202	27,280	26,048	24,655	24,130	32,626	25,980
10	Bing Organic	521	509	596	626	618	449	589	610	583	479	511	486	515
11	**SE Organic**	23,912	24,638	26,040	27,226	26,650	25,172	26,791	27,890	26,631	25,134	24,641	33,112	26,495
12	Kelkoo	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Pricerunner	2,951	3,308	3,411	3,908	3,773	3,352	3,899	3,834	3,336	3,070	2,977	3,212	2,743
14	**Price Services**	2,951	3,308	3,411	3,908	3,773	3,352	3,899	3,834	3,336	3,070	2,977	3,212	2,743
15	**Influencers**	9,991	16,568	15,147	15,459	14,536	14,246	17,497	18,745	18,350	19,001	18,029	18,736	11,839
16	Partner-ads	2,651	2,800	2,978	3,043	2,951	2,843	3,153	3,246	3,166	3,026	3,001	3,427	3,830
17	Adservice	0	0	0	0	0	0	0	0	0	0	0	0	0
18	**Affiliates**	2,651	2,800	2,978	3,043	2,951	2,843	3,153	3,246	3,166	3,026	3,001	3,427	3,830
19	**Social**	980	2,274	5,687	3,935	508	909	811	483	714	62	43	19	37
20	fleamarket.com	23	34	30	14	17	12	7	13	12	7	5	7	2
21	facebook marketplace	1,229	1,205	1,192	1,218	996	1,215	1,164	1,121	903	1,125	1,061	1,389	1,045
22	**Display**	1,252	1,239	1,222	1,232	1,013	1,227	1,171	1,134	915	1,132	1,066	1,396	1,047
23	E-mail subsciption	71,405	70,657	48,258	47,859	46,388	38,080	59,706	64,057	53,617	61,077	54,381	57,332	52,201
24	E-mail selected members	25,091	33,516	27,291	15,681	21,022	22,035	22,267	14,814	31,916	21,179	34,073	21,682	18,126
25	**Email**	96,496	104,173	75,549	63,540	67,410	60,115	81,973	78,871	85,533	82,256	88,454	79,014	70,327
26	Other	2,925	2,767	2,770	3,883	5,593	6,148	2,590	2,912	2,421	3,098	3,417	3,417	2,752
27	**Total**	238,288	279,999	256,412	244,198	249,630	231,999	265,559	263,260	265,169	256,955	251,457	233,930	215,802
28	*Per device*													
29	desktop	82,728	93,594	82,465	80,551	82,499	61,454	82,410	84,725	86,370	78,120	79,542	68,717	65,464
30	tablet	54,701	62,199	55,627	51,915	53,298	49,224	57,753	56,361	56,729	53,404	52,626	50,755	46,486
31	mobile	112,295	136,233	122,830	120,269	118,164	128,198	131,283	135,191	135,776	138,050	124,137	122,517	111,648
32	**Total 2**	249,724	292,026	260,922	252,735	253,961	238,876	271,446	276,277	278,875	269,574	256,305	241,989	223,598
33														
34														

	Channels (Traffic)	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31	W 32	W 33	W 34	W 35	W 36
3	Direct (organic)	26,904	27,749	29,851	31,453	31,321	30,684	31,245	36,861	32,226	31,216	28,277	32,063	32,967
4	Direct (paid)	17,617	17,104	17,565	18,339	17,473	17,879	17,713	21,069	19,162	18,177	15,927	18,518	18,052
5	**Direct**	44,521	44,853	47,416	49,792	48,794	48,563	48,958	57,930	51,388	49,393	44,204	50,581	51,019
6	Google Adwords	65,520	61,700	67,103	64,946	63,085	99,025	69,244	76,360	68,621	54,115	48,496	64,452	61,164
7	Bing Ads	1,354	1,340	1,279	1,370	1,182	1,243	1,214	1,251	1,310	1,420	1,226	1,227	1,206
8	**SE PPC**	66,874	63,040	68,382	66,316	64,267	100,268	70,458	77,611	69,931	55,535	49,722	65,679	62,370
9	Google Organic	23,262	22,153	22,540	23,994	22,499	23,225	23,542	26,259	24,140	21,635	19,447	22,082	23,344
10	Bing Organic	516	547	550	587	469	480	470	547	522	495	468	485	508
11	**SE Organic**	23,778	22,700	23,090	24,581	22,968	23,705	24,012	26,806	24,662	22,130	19,915	22,567	23,852
12	Kelkoo	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Pricerunner	2,998	2,907	2,647	2,769	2,559	2,518	2,666	2,849	2,786	2,470	2,159	2,042	1,933
14	**Price Services**	2,998	2,907	2,647	2,769	2,559	2,518	2,666	2,849	2,786	2,470	2,159	2,042	1,933
15	**Influencers**	12,179	22,121	22,473	22,164	22,619	31,846	26,674	24,021	19,460	15,306	15,117	18,764	17,673
16	Partner-ads	3,148	2,750	2,621	2,596	2,651	2,884	2,908	3,519	3,155	2,792	2,541	2,444	2,624
17	Adservice	0	0	0	0	0	0	0	0	0	0	0	0	0
18	**Affiliates**	3,148	2,750	2,621	2,596	2,651	2,884	2,908	3,519	3,155	2,792	2,541	2,444	2,624
19	**Social**	1,013	41	291	104	135	1,859	20	22	18	681	9	1,480	1,786
20	fleamarket.com	1	1	0	0	1	0	0	0	0	0	1	0	0
21	facebook marketplace	951	1,012	1,052	1,226	1,094	997	1,305	1,254	997	809	829	813	683
22	**Display**	952	1,013	1,052	1,226	1,095	997	1,305	1,254	997	809	830	813	683
23	E-mail subsciption	54,206	51,970	42,215	43,492	65,408	58,627	66,339	52,242	76,044	65,302	70,788	62,148	63,304
24	E-mail selected members	19,043	19,806	26,590	22,253	20,536	21,657	31,189	21,753	22,601	32,084	22,174	23,250	18,466
25	**Email**	73,249	71,776	68,805	65,745	85,944	80,284	97,528	73,995	98,645	97,386	92,962	85,398	81,770
26	Other	2,062	1,930	2,199	1,336	1,050	1,935	1,587	3,551	1,401	1,306	1,209	2,126	5,092
27	**Total**	230,774	233,131	238,976	236,629	252,082	294,859	276,116	271,558	272,443	247,808	228,668	251,894	248,802
28	*Per device*													
29	desktop	68,277	69,194	71,243	71,128	68,646	74,530	69,554	74,402	76,944	74,800	66,178	72,168	70,773
30	tablet	50,084	46,786	47,181	47,697	49,586	69,098	56,431	54,900	54,453	51,945	45,651	49,538	48,490
31	mobile	127,447	123,655	127,800	126,984	141,973	157,948	156,641	150,324	146,502	128,720	126,669	138,258	134,333
32	**Total 2**	245,808	239,635	246,224	245,809	260,205	301,576	282,626	279,626	277,899	255,465	238,498	259,964	253,596
33														
34														

35	Channels (Conversions)	W 11	W 12	W 13	W 14	W 15	W 16	W 17	W 18	W 19	W 20	W 21	W 22	W 23
36	Direct (organic)	1.88%	1.36%	1.85%	2.22%	2.21%	2.02%	2.47%	2.50%	2.31%	1.98%	2.33%	2.62%	2.31%
37	Direct (paid)	4.41%	4.22%	4.71%	4.93%	5.26%	4.02%	5.10%	5.12%	4.96%	4.23%	5.08%	4.05%	4.20%
38	**Direct**	2.80%	2.14%	2.77%	3.19%	3.30%	2.76%	3.45%	3.45%	3.26%	2.80%	3.30%	2.94%	2.93%
39	Google Adwords	2.17%	2.10%	2.26%	2.44%	2.28%	1.87%	2.38%	2.56%	2.62%	2.08%	2.33%	2.50%	2.09%
40	Bing Ads	3.12%	3.29%	3.07%	3.14%	3.48%	2.75%	3.25%	3.31%	3.24%	3.18%	3.55%	3.42%	3.27%
41	**SE PPC**	2.20%	2.14%	2.28%	2.46%	2.31%	1.88%	2.40%	2.58%	2.63%	2.11%	2.36%	2.52%	2.12%

35	Channels (Conversions)	W 24	W 25	W 26	W 27	W 28	W 29	W 30	W 31	W 32	W 33	W 34	W 35	W 36
36	Direct (organic)	2.41%	2.55%	2.45%	2.23%	2.13%	2.28%	2.45%	2.27%	2.30%	1.97%	2.15%	2.10%	1.86%
37	Direct (paid)	4.13%	4.54%	4.67%	4.15%	4.05%	4.11%	4.51%	4.16%	4.03%	3.78%	3.90%	3.97%	4.07%
38	**Direct**	3.09%	3.31%	3.27%	2.94%	2.82%	2.95%	3.19%	2.96%	2.95%	2.64%	2.78%	2.78%	2.64%
39	Google Adwords	2.15%	2.36%	2.20%	2.28%	2.08%	1.36%	1.94%	2.01%	2.27%	2.13%	2.33%	2.36%	2.34%
40	Bing Ads	3.36%	3.73%	3.87%	4.09%	3.64%	3.74%	4.12%	3.76%	3.40%	2.96%	3.34%	3.75%	3.57%
41	**SE PPC**	2.18%	2.39%	2.23%	2.32%	2.11%	1.39%	1.98%	2.04%	2.29%	2.15%	2.36%	2.39%	2.36%

Weekly | Long term data | 2019 Averages | Responsibilities | Raw Data

Appendix 2: Survey questions

Survey questions had the possible answers ranging from 1-5 with different values for each question, which are shown in the parentheses below each question.

Survey questions at the beginning:

How would you rate your data analysis experience?
(1 none | 2 basic | 3 moderate | 4 advanced | 5 expert)

How would you rate your Excel proficiency?
(1 none | 2 basic | 3 moderate | 4 advanced | 5 expert)

How do you feel right now?
(1 very calm | 2 calm | 3 moderate | 4 overwhelmed | 5 very overwhelmed)

Case questions:

1. Which two channels do you think are the most successful overall?
   Please choose 2 channels.

   1.1 Please explain your choices.

2. Some channels are reaching their potential. Which of these channels are the most saturated? Please choose 2 channels.

   2.1 Please explain your choices.

3. How do you expect the platform distribution will change in a year from now?
   Strategically, which platform would you focus on developing?
   (Desktop-Tablet-Mobile) Please choose a platform.

   3.1 Please explain your choice.

4. Which channel would you choose as the most suitable for a special promotion to boost short term visits? Please choose 2 channels.

   4.1 Please explain your choices.

5. Based on the information in sheet 4 (Responsibilities), to which channel(s) would you allocate additional work hours? Please choose 2 channels.

   5.1 Please explain your choices.

Survey questions at the end:

What is your gender?
(Female | Male | Prefer not to say)

How old are you?
(0-18 | 18-25 | 26-35 | 36-45 | 46+)

How do you feel right now?
(1 very calm | 2 calm | 3 moderate | 4 overwhelmed | 5 very overwhelmed)

How difficult did you find this exercise?
(1 not at all difficult | 2 slightly difficult | 3 moderately difficult | 4 quite difficult | 5 extremely difficult)

How complex was the data presented in the exercise?
(1 not at all complex | 2 slightly complex | 3 moderately complex | 4 quite complex | 5 extremely complex)

How clear did you find the instructions of the exercise?
(1 not at all clear | 2 slightly clear | 3 moderately clear | 4 quite clear | 5 extremely clear)

How focused were you during the exercise?
(1 not at all focused | 2 slightly focused | 3 moderately focused | 4 quite focused | 5 extremely focused)