Estonian Business School

Department of Marketing and Communication

DATA VISUALISATION IN WEBSITE USER BEHAVIOUR ANALYSIS IN THE CONTEXT OF DIGITAL ANALYTICS PLATFORMS

Bachelor's Thesis

Silver Ringvee

Supervisor: Sami Soinio

Tallinn 2018

I herewith declare that I have written the Bachelor Thesis independently. References have been indicated for the all publications, claims, opinions and different sources by other authors.

May 2, 2018 .... / Silver Ringvee

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

Today, more and more companies have realised that making decisions based on data is the key to success. Besides making it easier to gain new insights about the business and discover new business opportunities, data also helps lowering costs by removing the need for a trial and error method in many cases where the result is foreseeable using the tools of data analysis. Most companies are collecting some type of data, but since the rise of online media, it has never been easier nor more popular. The field of business analytics is becoming progressively more popular, and new software is constantly being built. The amount of data that is being collected online is enormous and growing rapidly, thus the way we analyse it is becoming increasingly more important. Besides making it easier to consume for others, the right visualisation techniques make data analysis both more accessible and faster,

The goal of this exploratory research is to study different methods of visualising user behavioural data in a digital analytics platforms context. The visualisation methods included in this research are those found in most popular user behaviour analysis tools (table 1). While there are more innovative, often artificial-intelligence-based methods, due to the scope of bachelor's thesis, only the popular data types are included, as most digital analysts are already familiar with these methods. Such digital analytics platforms, and therefore this research is aiming for an audience coming from either digital marketing or analytics background. The primary purpose of this paper is to study the relationship of data visualisation aesthetics and raw data by measuring the perceived user experience satisfaction level in a digital analytics platform. Therefore, the research question is: how significant an impact does data visualisation have on the daily work of digital analysts and online marketers?

The author of this research has been working with many digital analytics platforms for more than four years, and during that time, he has seen users with a multitude of different issues related to using those platforms. Many of those problems have been in one way or another related to data visualisations. Usually, visualising more general digital analytics data such as visitor numbers, goal completions, and revenue is done well, using lines and other basic charts. The real problem, though, seems to occur when user behavioural data comes into play, as many of it cannot be represented using a simple line or bar chart, and a more creative solution is required. This research is going to cover the popular types of user behavioural data used in digital analytics, how common tools have solved visualisation part and what problems do the users have related to those visualisations.

In this research, different approaches to visualising the metrics and dimensions that help us understand the behaviour of website users were studied. The methods that will be used in this research include reviewing the literature and previous studies as well as conducting a series of Think-aloud observations amongst people using this sort of data in their daily work (online marketers and analysts) to see which visualisation options are the most popular and which ones are rarely used. The participants will be guided to use tools that have all these features in place. Furthermore, the observations will find out which data visualisation methods help them to find what is essential the fastest, e.g. comparing getting insights from heatmaps to working with raw numbers.

This research will most directly affect a specific software company that is building a digital analytics platform that focuses on website user behavioural data. The author of this research is a founder of the company and therefore has the power to make such decisions. The company is called Reflective Data, and is currently serving around hundred clients across the world, collecting data from millions of users every month. Visualising this kind of data can be difficult, therefore finding an optimal solution is crucial. Furthermore, a data-driven approach to such vital questions is a must, especially for a company that is encouraging others to make their decisions based on data-based visual representations. As this software has already made its way into a daily routine of many users, they will be directly affected by any changes that will appear in the interface.

1. LITERATURE REVIEW

When it comes to literature and previous studies, the author will be covering those done in the areas of data visualisation, digital analytics, and online user behavioural analysis. Although the history of data visualisation dates back to 17th century (Few, 2013, p. 35), the fields of digital analytics and website user behaviour analysis are rather new. Nevertheless, they are both rapidly growing and so is the amount of research that focuses on the related topics. The author has searched from EBSCO, Google Scholar, ScienceDirect and IEEE Xplore with keywords like “analysts user experience”, “user behaviour analysis”, “visualising in digital analytics”, and “data visualisation” and no previous research on this topic was found. There is much research on measuring end-user experience on a website, but no research on measuring the analysts’ user experience in a digital analytics platform.

The literature review of this research is divided into four parts. The first part gives an overview of data visualisation and its history in general. The second part describes the concept of visual hierarchy and its most essential elements. The third part introduces some of the most popular user behaviour analysis platforms. The fourth part covers three different common types of user behavioural data.

1.1 Data Visualisation

Data visualisation papers are often more informal than in other domains (Elmqvist, 2016). This manifests in several ways: (1) the tone is often more informal than in other domains, (2) many (although not all) information visualisation papers have a more engineering-centric feel, and (3) human-subjected experiments regularly include a smaller number of participants than in the social sciences or medicine (on the order of 10-25 rather than 100-250), minimal individual differences (compared to effect sizes), and a large number of trials.

While the process of arranging data into tables (columns and rows) has been around since the 2nd century, and the graphical representation dates back to 17th century (Few, 2013, p. 35), the real popularity of data visualisation started in 1999. This was when a new research field was coined "information visualisation" (Few, 2013, p. 36) in the 1999 book, *Readings in Information visualisation: Using Vision to Think*. In this book, Stuart Card, Jock Mackinlay, and Ben Shneiderman for the first time combined the academic research that had been done earlier into a single book. That was the first time when this kind of research in the field of data visualisation was made publicly accessible beyond the walls of academia (Card et al., 1999).

Visualisations make a big part of most digital analytics platforms and just like other everyday things, general computer interfaces, and visualisations can be designed either well or poorly (Norman, 2013). Poor design causes frustration, waste of time, and is generally bad for both the user and the owners of the software. At the same time, a good design can save time, make users happy and is undoubtedly good for the business (Norman, 2013). A good design is human-centric, putting the end user of a product or service first, everything else, including the looks, cost, and engineers' opinions should come second. Such design methodology is known as The Human-Centered Design Process (Norman, 2013), this term was first used and thoroughly described by Donald Norman in 1988 in his book *Psychology of everyday things*.

In this research, the author was not studying data visualisation in general but in a context of digital analytics platforms. Stephen Few said in his article (2009, p. 2) that data exploration requires the ability to quickly spot and examine everything that is potentially meaningful in a dataset with relative ease. The tool that we use should allow our minds to remain engaged in thinking about the data without distraction. In other words, ideally, the tools should disappear from our awareness so our thoughts can remain focused on understanding the data, without any distraction from the mechanics of working with the tool. Based on that knowledge, one can conclude that data visualisation in a tool should be clear, easy to read and the usability should be similar to the tools that have become an industry standard (to reduce the learning curve), e.g. Microsoft Excel or Google Analytics.

1.2 The Concept of Visual Hierarchy

The concept of visual hierarchy refers to the arrangement and presentation of elements in a way that implies importance, often achieved by using visual contrast (GIS Dictionary, 2014). This concept is based on Gestalt psychological theory from the early 20th century in Germany. The Gestalt's theory proposes that the human brain has innate organising inclinations which structure individual elements, shapes or forms into a coherent, organised whole (Jackson, 2008, p. 63-69). As data visualisation consists of different visuals, and is consumed using vision, the concept of visual hierarchy is highly relevant to this research.

It is known that human brain disassociates objects from one each other mostly based on the differences between their physical characteristics. These characteristics can be divided into four main categories: colour, size, alignment, and character (Feldsted, 1950). All of these four categories are relevant to data visualisation, even tables and other basic forms of visualisations can be improved by changing these variables. Not to mention the more advanced visualisations like heatmaps and word clouds.

As with most visuals, colours have a very significant role in data visualisation. We need to consider three separate effects during colour selection: colour distance, linear separation, and colour category (Healey, 1996, p. 1). Healey's research analysed the ability of thirty-eight observers to find a colour target in displays that contained differently coloured background elements. Their results showed that their method could be used to select a group of colours that will provide a proper differentiation between data elements during data visualisation. In the list of tools that are studied in this research are some with very strong focus on colours and others with somewhat plain looks.

Using the principles of visual hierarchy in web user experience design has become a common practice. Such techniques are being used to help viewers find a point of entry into the page (searching) as well as viewer's behaviour after finding such an entry point (scanning) (Djamasbi et al., 2011, p. 331-340). In the context of digital analytics platform, a visual hierarchy can have an effect on which reports and visualisations the user starts to view and what information they find the first or will remember the best.

1.3 Digital Analytics Platforms With the Primary Focus on User Behaviour

The measurability of digital media has been heralded as one of its most significant benefits compared with other media since the mid-1990s when internet marketing, as it was known then, first started to be deployed (Chaffey, Patron, 2012, p. 1). As the popularity of digital media started growing, people soon realised that looking at the log files they could learn a lot about the consumption of their content. Moreover, by adding some metadata to those log files, they also started to learn more about the website visitors and their behaviour. In 1994, the first commercial web analytics vendor, I/PRO Corp, was launched. After that, we have seen general digital analytics tools like WebTrends (1995), Omniture (2002) and Google Analytics (2005) launch and soon after gaining a lot of popularity (Chaffey, Patron, 2012, p. 2).

General digital analytics platforms mentioned in the previous paragraph provide a good overview of website's key performance indicators and other general metrics (Minkara, 2017, p. 7). On the other hand, these tools might not be the best when the goal is understanding the behaviour of the users and what triggers might be behind specific actions they take. The second kind of information, though, is highly useful when optimising for website's better user experience (Minkara, 2017, p. 8). Therefore, the tools that specialise in website user behaviour analysis have become more popular year by year. Some of the most popular digital analytics platform specialising in the user behaviour analysis are Hotjar, Mouseflow, Decibel Insight, Clicktale and

Inspectlet (table 1). While none of them is as famous as Google Analytics which is used by 64.8% of the top 1 million websites (table 1), Hotjar is already quite popular and is used by 3.1% of the top 1 million websites, followed by Mouseflow (0.3% of 1M).

Table 1. Digital analytics platforms with the primary focus on user behaviour, ordered by popularity. Of top 1 million websites. Data source: BuiltWith (online database) https://trends.builtwith.com/analytics/

Software name	Active installs (of top 1M)
Google Analytics	64.8%
Hotjar	3.1%
Mouseflow	0.3%
Clicktale	0.1%
Decibel Insight	<0.1%

Besides the top five listed in table 1, this research also touches platforms known as Reflective Data and Inspectlet. The most popular of all of them, Google Analytics, is, based on their official documentation, focusing on being a good fit for general-purpose digital analytics (Google, 2018). Hotjar is marketing themselves as an easy-to-use all-in-one user behaviour analysis platform (Hotjar, 2018). Mouseflow is a feature-packed user behaviour analysis software focusing on larger clients (Mouseflow, 2018). Clicktale is an enterprise-focused experience analytics platform (Clicktale, 2018). Much like Clicktale, Decibel Insight is also an enterprise-focused experience and user behaviour analysis software (Decibel Insight, 2018). Inspectlet is a user behaviour analysis tool focusing on session replays and heatmaps (Inspectlet, 2018) and Reflective Data is a new platform focusing on general user behaviour and engagement analysis (Reflective Data, 2018).

1.4 Types of User Behavioural Data

While the variety of different types of website user behavioural data that could be collected using modern tracking technologies is rather large, this research is going to focus on three key types of website user behavioural data. Those three can be tracked by default in most user behaviour focused digital analytics platforms such as Hotjar, Clicktale and Decibel Insight (table 1 and table 2). Therefore, users are familiar with them and should know how to use them in their everyday work. While on-site polls and session replays are both available in 3 out of 5 most used platforms (table 2), only the first one is included in this research. The reason for excluding session recordings is that there is no other way to visualise a video recording and the author of this research has more experience with on-site polls. In the following paragraphs, there is an overview of each data type that will be covered in the research.

Table 2. Top features of digital analytics platforms with the primary focus on user behaviour. Data source: platforms' websites

Feature/data type	Available in top 5 tools
Heatmaps	5
Form analytics	4
On-site polls	3
Session replays	3
Funnel visualisations	2
Recruiting test users	1

1.4.1 Mouse Movement, Clicks and Scrolling

Clicks, scrolling, key presses, and screen touches are the main ways how people interact with websites (Liikkanen, 2017, p. 52). While there are other possible approaches like tables, bar charts or even pie charts, all of the top analytics platforms are using some form of heatmaps to visualise mouse-related user behavioural data (Hotjar, Clicktale, Decibel Insight). Heatmap refers to data visualisation technique where colours or tones of colours represent the power of a specific metric (Wilkinson, 2009, p. 3). Heatmaps could be drawn on a matrix, on a geographical map or on some other type of visual (a picture of a house, website etc.) (Wilkinson, 2009, p. 3). In the online behavioural analysis, heatmaps are mostly used to analyse mouse movement, mouse clicks, and scroll depth (table 2). More advanced tools will also collect the time that was spent at specific areas of the website combining information about both scroll depth and mouse movement, usually called attention heatmap (Bonggun et al., 2017, p. 8).

Although the name “heatmaps” and the form as we know them today was originally trademarked by the software developer Cormac Kinney in 1991 (Wilkinson, 2009), similar plots have been used for various purposes for more than a century (Wilkinson, 2009). Kinney initially used heatmaps to describe a 2D display depicting financial market information (Wilkinson, 2009), since then we can see them being used in a wide variety of different scenarios and use cases, online user behaviour analysis being one of them. Since the very close connection between online marketing and digital analytics in the past decade, we can see the popularity of heatmaps continually growing (Bojko, 2009, p. 3), the key reasons for that are the appealing look and the (often false) feeling that they provide considerable amount of information in just a quick overview (Bojko, 2009, p. 10), this problem is well researched by Agnieszka Bojko and the following quote describes it very accurately.

> We are often blinded by the attractiveness and apparent intuitiveness of heatmaps, and so we often do not realise how much information in addition to the visualisation itself is necessary to understand the heatmap fully and correctly interpret the data it represents. In other words, the biggest danger involved in creating and reading heatmaps is that we are often unaware of what we do not know, and thus we do not look or ask for the missing information. (Bojko, 2009, p. 10)

In general, online user behavioural heatmaps are generated based on mouse movement, clicks and scrolling. To make things more advanced, analysts and scientist are also using technology that is called eye tracking (Granka et al., 2004, p. 1). An eye tracker is a camera lens placed closely next to a computer’s display and it then accurately tracks where a visitor is looking. Although, in past years such technology is not as expensive as it used to be (Duchowski, 2003, p. 45), the lack of knowledge how to properly install and use them makes them a rather unpopular choice amongst both the online marketers and digital analysts, not to mention the fact this experiment requires a lab environment and the samples are much lower when compared collecting data remotely (Duchowski, 2003, p. 97). As there is a tolerable correlation between gaze and mouse movement (Liebling et al., 2014), heatmaps generated based on data collected on mouse interactions are thought to be an easy and fast way for getting an overview of the website's usage as they are rather simple to both set up and read. Besides, one does not have to be an expert to tell which areas got the most attention or which areas seem to stay unnoticed (Bitkulova, 2017, p. 34).

1.4.2 Form Analytics

Form Analytics is a method of analysing how people interact with the online forms. The key metrics of this dataset are form impressions (number of visitors who saw the form), interactions (people who interacted with the form, most commonly it is the first click), conversions (people who filled in and then successfully submitted the form). To make analysis more meaningful, metrics like exits on a specific field, time spent on field and corrections could also be collected. The fact that on most websites a form is the separator between a product and the client (Wroblewski, 2008, p. 4) makes them the most important part of the website (end of the funnel, checkout page, quote form etc.), knowing how people interact with them is a crucial part of optimising the user experience and website in general.

A well-designed form can increase both the number of users that start filling it and the information they decide to enter, not to mention those who eventually submit the form. As the submission of a web form is often directly related to how much money a company makes, getting this part of the website right is crucial. It is not rare to see a form completion rate go up by 10-40 percent after a data-driven re-design (Wroblewski, 2008, p. 16). Most website user behaviour analysis tools have some sort of form analytics feature built in. Some of them are rather primitive while others provide a great variety of advanced features such as combining multiple forms into a funnel. One of such tools is Formisimo (Brooks, 2017).

1.4.3 On-site Polls

On-Site Polls are small questionnaires that pop up on the website based on a pre-set ruleset. Usually, after a specific action from a website user, e.g. exit intent or a certain scroll depth (Vaht, 2015, p. 14). In general, these polls are not longer than two questions, most commonly it is just one (Vaht, 2015, p. 14). Both open and closed-ended questions are being used. Here are some examples of the questions that might be asked on a website: What holds you back from buying this product? How likely are you to recommend [PRODUCT / BRAND / SERVICE] to your friends or colleagues (Darmanin, 2016)? The latter gives us the net promoter score (NPS), which is a well-known indicator of customer satisfaction, this is a strong predictor of a company's ability to drive sustainable profits and growth in the future (Mattox, 2013). On-site polls can be very insightful because instead of guessing based on analytical numbers, an analyst could ask directly from people what concerns they are having. The problem with polls is that an analyst often ends up having a lot of qualitative data that has to be coded for further analysis. By using creative data-visualisation techniques, such as word clouds and word frequency, it is still possible to find some valuable insights, without any coding (Cui et al., 2010).

A big part of the data that on-site polls provide is qualitative (Laja, 2017). Quantifying qualitative data has been rather widely researched (Forrest et al., 1980) and in the past years with the rising popularity of machine learning and artificial intelligence new solutions have been developed (Honkela, 2016). Today, word clouds are very popular for visualising website related data, for example for giving a broad content overview (Cui et al., 2010). When it comes to visualising the data from on-site polls, besides word clouds also simple word frequency bar charts and word counts (counts of individual words) are being used (Hotjar, 2017). Usually, punctuation, pronouns and other common words are excluded, as they do not have much meaning in that context.

Some great research has been done related to word clouds by the scientists from Hong Kong University of Science and Technology in cooperation with the IBM China Research Lab. Together, they introduced a visualisation method that combines word clouds with a trend chart to illustrate temporal content evolutions in a specific set of documents (Cui et al., 2010). Furthermore, they used a trend chart to encode the overall semantic evolution of document content over time. This kind of solution would be extremely useful for on-site polls that run for a longer period of time as they would help to detect any long-term changes in overall trends (Cui et al., 2010).

1.4.4 Metadata

Based on the three categories of user behavioural data mentioned above, the data that is being collected on websites contains several different metrics and dimensions. For forms, those are impressions, interactions, form submissions and time spent on each field (Bolton, 2017). For heatmaps, the information about each click, mouse movement and scroll depth is needed (Hotjar, 2017). For the polls, the data consists of answers to one or more open and closed-ended questions (Vaht, 2015, p. 14). To put this data into context, a good amount of metadata has to be collected. Metadata is data that provides information about other data (Merriam Webster, 2018), the goal is to put the main data better into context. While the variety of different metadata that could be collected is virtually limitless, most of the tools are collecting data about the website (current subpage, referring website, UTM parameters etc.) and user (device, browser, geographical location, language, timezone etc.) (Hotjar, 2018). Some dimensions are used for building general reports, and others are mostly good for building custom segments that can be used for further analysis, e.g. comparing mobile users to desktop users. In general, it is good to track as many dimensions as possible so they could be later used in analysing website user behaviour in greater detail (Jeon et al., 2008), although, since GDPR, users should be careful with everything that could be interpreted as personal information (Bouquet, 2017).

Many tools such as Hotjar and Clicktale let their users collect and store more advanced data that is specific to their website in custom dimension (Hotjar, 2018). Due to the structure of relational databases that are being used for storing the data, the number of custom dimensions is usually limited to 5-25 slots (Google Analytics, 2018). Custom dimensions are good for storing information specific to the user (user ID, age, gender etc.) and their actions (logged in/out, number of products purchased etc.) (Google Analytics, 2018). The information from a custom dimension is almost never used in default reports, and analysts instead leverage them when building custom segments. For example, to compare the behaviour of those who have made a purchase and those who have not (Google Analytics, 2018).

2. RESEARCH METHODOLOGY

When it comes to finding the right methodology, the author took into account the goal of the research, which is to study the solutions for visualising user behavioural data in the context of digital analytics platform and to answer the central question of this research: how significant an impact does data visualisation have on the daily work of digital analysts and online marketers? These two groups were heavily involved in the process of answering this question. The segment of internet users that was the subject of this research commonly already has a rather good overview of the different visualisation options, they are familiar with different charts and can learn them faster than average internet user. Therefore, the visualisations that were studied are not necessarily as simple as possible but rather as informative as possible.

2.1 Research Methods

The chosen methodology for collecting data was Think Aloud Protocol (or Talk Aloud Protocol). While this is a less-known method in general research, it has already been widely used to gather data in usability testing, psychology and in a variety of social sciences. Data visualisations are a significant component of a digital analytics platform's usability, and Pauline Genise stated (2002) that Think Aloud Protocol provides the results that are close to what is experienced by users, therefore this method considered to be suitable. While there are some variations of the Think Aloud method, using it in usability testing was introduced by Clayton Lewis at the time he was at IBM and is explained in *Task-Centered User Interface Design: A Practical Introduction* (Lewis, 1982). A known disadvantage of Think Aloud Protocol that the environment is not natural to the user (Genise, 2002). In this research, this was eliminated by letting the participants work in their normal working environment using the tools they are familiar with.

Think Aloud Protocol involves participants "thinking aloud" while they are doing a set of specified tasks (Jääskeläinen, 2010, p. 371). Participants are asked to say whatever comes into their mind while they are completing the tasks. Thoughts might include what they are looking at, thinking, doing, and feeling, as well as why they decided to do action X instead of action Y (Jääskeläinen, 2010, p. 371). This process gives observers insights into the participant's cognitive processes (rather than only their end product), to make the entire thought processes as explicit as possible during task performance (Kucan et al., 1997, p. 272). In a usability testing context, the observer was taking notes of what participants said and did, while doing so he especially took note of places where they encountered any difficulties. It is essential to let the participants work without being interfered or influenced in any way by the observers (Kucan et al., 1997, p. 273). The sessions are often audio- and video-recorded so that developers, analysts, and designers can go back and refer to what participants did and how they reacted. In this research both audio and the video was recorded. Generally, during studies based on the Think Aloud method, research subjects must spend several minutes (sometimes more than one hour) working their way through the given tasks (Johnstone et al., 2006, p. 4).

While choosing the right methodology, the following ones were also in consideration. Experimental research, where a more extensive test group would see different visualisations compared to the control over the time of using the platform without them knowing about the experiment. The differences in several behavioural trends and metrics between the control and the test group would have been studied. While running such experiment is possible, it would have taken a lot of time and resources, and the sample size would have been rather big in order to say for sure that a change in behaviour was indeed caused by the changes in the variation. This method was left aside due to not being the best match for the scope of this research.

Another option that was in the consideration included conducting an online survey consisting of both qualitative and quantitative questions. The sample would have consisted of both online marketers and digital analysts from the author's professional contacts as well as from a variety of groups and online forums on that topic. There would have been no geographical or demographical restrictions as the results should be applicable to anyone working in the industry, and most of the tools are being marketed globally. The reason for not using this method was that Thinking Aloud Protocol was specifically designed for gathering data in usability testing.

After making some thorough comparisons of each methodology, the Think Aloud Protocol seemed to be the most suitable in the context of this research, especially due to the fact that it was specifically designed for gathering data in usability testing in product design and development (Lewis, 1982). Despite to some criticism against the Think Aloud Protocol, especially the rather small sample size that one can collect and analyse with the resources available (Johnson, 1992, p. 71), and the fact that the results of such research heavily depend on analyst's skills (Charters, 2003), this still looked as the best method as the author has experience in using similar research methods from his professional work. Not only is it meant for collecting and analysing a lot of qualitative data, but it also has already proven to be successful in various usability testing research and is widely used in the industry (Genesis, 2002). While this method might not be the easiest to conduct and analyse, it provides the most useful insights and does not limit the actions/thoughts of the target group in any way, i.e. not giving them the list of options in a survey (Esch, 2013).

2.2 Data Collection Plan

In this research, the participants were using the common digital analytics platforms (Hotjar, Decibel Insight, Clicktale, Reflective Data, Mouseflow, Inspectlet) that have various user behaviour visualisations in place but all of them include three data types that are in the main focus of this research: heatmaps, form analytics and on-site polls. The participants were allowed to use other tools if that is something they would do in their everyday work. It is important though that those digital analytics tools have at least one of the common user behaviour data types in place.

While going through the protocol, the participants searched for insights and took other actions from their daily routine. The sessions took place in their normal working environment, using the tools they are normally using. While doing so, they were "thinking aloud", and the whole process was recorded. Later, the recordings were transcribed and analysed; this is the standard workflow for any Think Aloud Protocol based research (Jaspers et al., 2014). During the session, the observer was taking notes of what the participant said and did, especially at places where they encountered difficulty or something else noteworthy happened. Participants were encouraged to let the observer know if the information they see on the screen is valuable or rather useless. Furthermore, participants were advised to inform the observer of the quality and the accuracy of the insights they find, and both were rated on a five-point scale. As analysts tend to double-check the insights they find in behaviour analysis platforms by using raw numbers, they were asked to do the same in our sessions. While the observable was working, the observer did not interact with the observable in any way. The only allowed interaction was to remind the participant to talk, in case they were taking actions and not adding any comments. A set of questions was asked from the participant before or after the session; this gave a better overview of who the participants are as well as what is their professional background.

The location and environment for these Think Aloud Protocol sessions was not fixed, but it was important that the participants are able to work without any major distractions. Therefore, a cafe would not be a good location but an office meeting room would work fine. Due to the fact that a lot of the experts in the field are not located in Estonia (the location of the author), some of the sessions were held over video calling software. This did not limit the amount of data that was collected, with modern software it is possible to simultaneously record participants screen, face and voice. For the sessions that take place on site, a similar set of recording software was being used.

Besides collecting qualitative data, the Think Aloud method also allowed collecting a set of quantitative data. The latter was achieved by asking the participants to evaluate the insights each platform and feature based on their accuracy and quality, and both were rated on a five-point scale. The method of combining qualitative and quantitative data collection and analysis is known as mixed methods approach (Saunders et al., 2009, p. 152) and in this research, a parallel form of this approach was being used. This means that the two data types are being collected simultaneously (Saunders et al., 2009, p. 152).

2.3 Sample Selection

Because Think Aloud Protocol, by its nature, is rather labour-intensive, the quite small sample size involved in the research was six experts of digital analytics. Small sample group, however, does not necessarily indicate small data sets. As the research process is in-depth, small sample sizes can still provide valid insights and information. Nielson (1994), for example, suggested that sample sizes as small as five participants will provide sufficient information about the problem-solving behaviour.

Additionally, unlike large surveys or psychometric research projects, samples are almost never chosen at random for Think Aloud Protocols. Quite often those groups themselves are marginal when compared to the entire population (Johnstone et al., 2006, p. 5). All of the participants in this research were chosen to represent the target group that was being studied. In this research, only digital analysts and marketers with at least two years of experience were chosen. Another rule for participants was that they must be currently using at least one user behaviour analysis tool in their everyday work. Therefore, using this method with limited resources available, it was not important to chase a large number of participants but rather get as much information from every participant as possible. For example, Morse (1994, p. 225) suggested that in qualitative research, the sample size for interviews should be at least six.

There have been many think-aloud protocol based research studies with a sample size of 5-10, to give a few examples:

- Jaspers et al. 2004. The Think Aloud method: a guide to user interface design. Sample size 4 pediatric oncologists.
- Johnstone et al. 2006. Using the Think Aloud Method (Cognitive Labs) To Evaluate Test Design for Students with Disabilities and English Language Learners. Sample size 5-10 students per group.

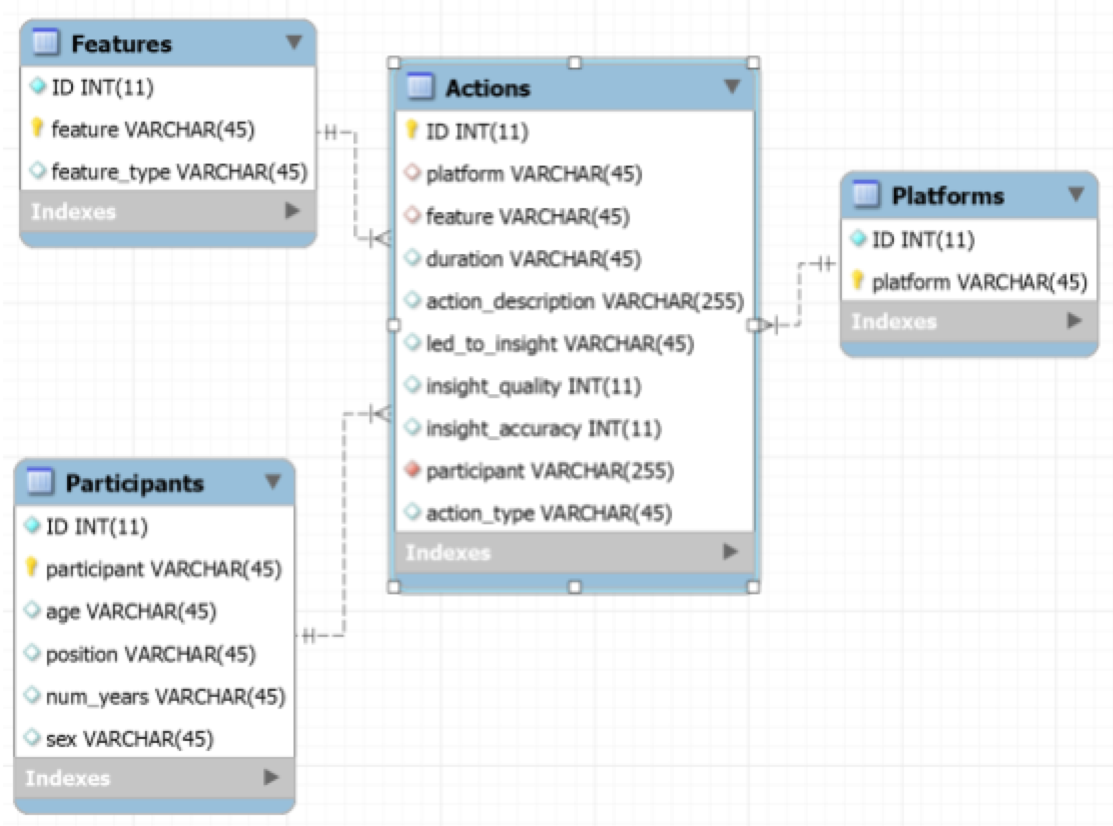
Taking into account what other researchers have concluded, in this research the sample size was set to be at least five participants, the final sample size ended up being six. To find the participants, the author contacted several online marketers and digital analysts from his professional contacts, making it a sample of convenience. Every participant had at least two years of experience in their field. This requirement should eliminate those who have experience with using only one tool and therefore might not be able to compare different visualisations.

2.4 Data Analysis

After all the Think Aloud Protocol sessions were done and recorded, they were transcribed for further analysis. To answer the main research question, the types of data that were used most often, how they were visualised and how the participants felt about different data types and visualisation solutions were taken into account. It was also important to figure out how significant an impact does the type of visualisation used have on the user and whether they struggle with finding what is important.

After sessions were transcribed, all user actions were stored in a SQL database (MySQL version 5.1.69). This allowed advanced queries for further analysis. The database was organised into four tables: Participants, Platforms, Features, and Actions. All tables in the database were connected based on at least one column (figure 1). Each action entry has ten attributes (columns) (1) index, (2) platform name, (3) feature name, (4) duration, in seconds, (5) action description, (6) led to insight, true or false, (7) insight quality, 1-5 point, (8) insight accuracy, 1-5 points, (9) participant name, and (10) action type. For each participant five attributes were stored in a table (1) index, (2) age, (3) sex, (4) number of years worked in the field, (5) position, job title. These data-points gave a strong foundation for further analysis, letting the analyst group, sort and filter the dataset based on any of the given columns.

Figure 1. Database schema. Source: Screenshot from MySQL Workbench.



Due to the type and structure of data that was being stored, a choice of using SQL as opposed NoSQL database was rather easy. Using a SQL database for storing user actions and other related information opened an opportunity for different types of data analysis, including grouping, sorting and filtering based on any stored value. Some examples of the possible queries include: (1) to see how many actions were recorded for each participant: SELECT participant, count(1) FROM Actions GROUP BY participant, (2) to see how many actions were recorded for each platform: SELECT platform, count(1) FROM Actions GROUP BY platform, (3) to see how much time was spent on each platform: SELECT platform, sum(duration) FROM Actions GROUP BY platform.

Furthermore, besides a SQL database, a set of popular data science tools (Stringfellow, 2018) was being used. The tools included Python (a programming language), Matplotlib for plotting, NumPy for mathematical functions and Pandas for data structuring and manipulations. Most of the analysis was done in a popular Jupyter Notebook data analysis environment. This allowed seamless connections with the database with fast queries and on-the-go visualisations with flexible plotting options. Additionally, this toolset enabled running a set of statistical models across the entire database. While using such tools is not a must in this type of research, it made processing and analysing data more structured and simpler.

The digital analytics platforms were compared based on multiple different criteriums: (1) number of times it was used, (2) the amount of time that was spent using them, (3) the amount of insights that they provided, (4) the quality and the (5) accuracy of the insights they provided. Furthermore, to give a better result, these metrics were combined several different ways, for example (1) the amount of time spent on finding an insight, (2) time spent by the quality of the insights found, and (3) time spent by the accuracy of the insights found. Similar criteriums were used for evaluating platform features and visualisation solutions.

Based on the data collected from the Think Aloud Protocol sessions combined with the toolset named before, the author of this research was able to analyse the results as well as further explore and study the influence of data visualisation of user behavioural data in the context of digital analytics platforms. As the participants were asked to double-check their findings with raw numbers, mostly in Google Analytics or in a custom database, the speed and quality of the two were also compared. The fact that participants used seven different platforms allowed searching for connections between the speed and quality of the insights and the different visualisation methods that the tools had implemented.

3. RESULTS, DISCUSSION AND RECOMMENDATIONS

This chapter introduces the data that was collected in Think Aloud Protocol sessions, transcribed and stored in the database, and later analysed using common tools of data science (Stringfellow, 2018). Chapter 3.1 gives a good overview of the data itself as well as connections between different behavioural trends and specific tools and visualisations. The chapter 3.2 further analyses the results introduced in 3.1 and gives some recommendations for Reflective Data and other companies in the same business. Lastly, chapter 3.3 gives an overview of some of the limitations of the current research.

3.1 Results of Think Aloud Protocol Sessions

Firstly, to put the results into better context, here are the key numbers related to the data collected from the Think Aloud Protocol sessions. This includes both the qualitative and quantitative data. In total, 6 participants took part in the sessions, five men and one woman. The average age of the participant was 35.3 years, the oldest of them being 42 and the youngest was 29 years old. They had between 3 and 16 years of experience with digital analytics or online marketing, on average 7 years of experience. During the sessions, 7 different digital analytics platforms were used: Hotjar, Clicktale, Reflective Data, Inspectlet, Mouseflow, Decibel Insight, and Google Analytics. A total of 306 actions were stored in a SQL database, the sum of the durations of all actions was 901 minutes or 15 hours. While working, the participants found 142 insights with different quality and accuracy metrics. The average quality of the insight (on a five-point scale) was 2.87, and the average accuracy was 3.29, these ranks were given by the participants during the sessions.

To follow the common practice of data science (Kluyver et al., 2016), the author has included a database query that was used to create a table or a graphic above each of them in the caption area.

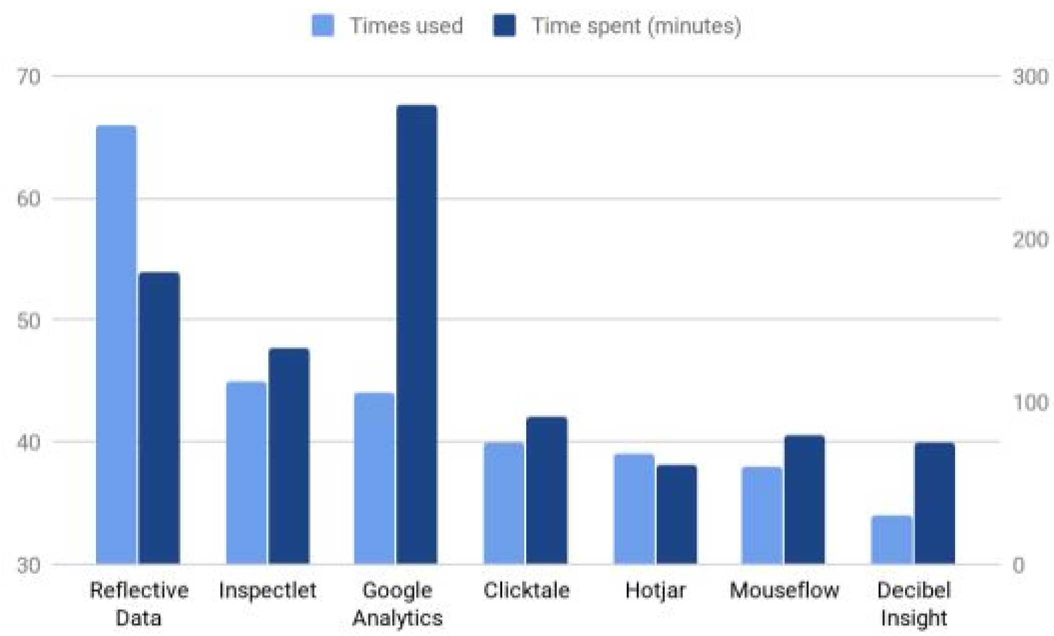
Comparison of the platforms provided several interesting trends. First of them is that the usage of the platforms was quite linear with no significant outliers (table 3).

Table 3. A number of times a platform was used by a participant and the amount of time that was spent using each platform. Source: Composed by the author. Query: SELECT platform, count(1) AS "times used", sum(duration) AS duration FROM Actions GROUP BY platform ORDER BY count(1) DESC

Platform	Times used	Time spent (minutes)
Reflective Data	66	180
Inspectlet	45	133
Google Analytics	44	282
Clicktale	40	91
Hotjar	39	61
Mouseflow	38	79
Decibel Insight	34	75

The amount of time spent using a specific platform (table 3) did not directly match with the number of times they were used (chart 1). That means that using one tool took longer than another. As can be seen from the chart 1, even though Reflective Data was used 40% more often than Google Analytics, the latter took 44.2% more of the users time. A significant difference between times used and time spent also occurred for Decibel Insight which seems to take a lot of user’s time when compared to other tools.

Chart 1. Difference between times a platform was used (left axis) versus the amount of time spent using them (right axis). Source: Composed by the author. Query from table 3



When comparing the insights each platform provided, several criteria were used in the evaluation process. Three key metrics were (1) amount of insights found, (2) quality of the insights found, and (3) accuracy of the insights found. The first metric was looked at based on two different dimensions: (1) amount of times the platform was used and (2) duration of platform usage. Due to the type of data they were holding, the latter two metrics were only compared to the duration dimension. Combining these metrics and dimensions, four different ratios were calculated (table 4).

Table 4. Platform comparison metrics. The number of insights found divided by the duration of time that was spent using this platform, insights found divided by the number of times a platform was used, average quality of the insights found divided by the duration that was spent using this platform, the average accuracy of the insights found divided by the duration that was spent using this platform. Highest values marked with green, lowest with red. Source: Composed by the author. Query: SELECT platform AS p, sum(duration) AS "time spent", count(1) AS "times used", (SELECT count(1) FROM Actions WHERE led_to_insight="true" AND platform=p) AS insights, (SELECT avg(insight_quality) FROM Actions WHERE led_to_insight="true" AND platform=p) AS insight_quality, (SELECT avg(insight_accuracy) FROM Actions WHERE led_to_insight="true" AND platform=p) AS insight_accuracy FROM Actions GROUP BY platform ORDER BY count(1) DESC

Platform	insights found by time spent	insights found by times used	insight quality by time spent	insight accuracy by time spent
Google Analytics	7.80%	50.00%	0.92%	1.13%
Reflective Data	17.22%	46.97%	1.61%	2.01%
Inspectlet	15.04%	44.44%	2.41%	2.33%
Clicktale	16.48%	37.50%	3.00%	3.88%
Mouseflow	26.58%	55.26%	3.92%	4.40%
Decibel Insight	28.00%	61.76%	3.75%	3.87%
Hotjar	19.67%	30.77%	4.23%	5.05%

As can be seen in table 4, Decibel Insight has the best insights to duration ratio as well as insights to times used ratio. This means that the platform is easy to use and shows the information the user is looking for. When we add the metrics for insight quality and accuracy, Hotjar is the best. This means that Hotjar can provide high-value insights the fastest and most probably indicates clear interface and meaningful visualisations. Google Analytics, on the other hand, performed rather poorly in all the key metrics. While it can provide good insights, working with raw numbers and basic visualisations takes a lot more time.

When comparing the insights each feature or data type provided, similar metrics were taken into account. Based on these metrics, four key ratios were calculated: (1) insights by duration, (2) insights by times used, (3) quality by duration, and (4) accuracy by duration. These four ratios can be seen in table 5.

Table 5. Feature comparison metrics. The number of insights found divided by the duration of time that was spent using this feature, insights found divided by the number of times a feature was used, average quality of the insights found divided by the duration that was spent using this feature, the average accuracy of the insights found divided by the duration that was spent using this feature. Highest values marked with green, lowest with red. Source: Composed by the author. Query: SELECT feature AS f, sum(duration) AS "time spent", count(1) AS "times used", (SELECT count(1) FROM Actions WHERE led_to_insight="true" AND feature=f) AS insights, (SELECT avg(insight_quality) FROM Actions WHERE led_to_insight="true" AND feature=f) AS insight_quality, (SELECT avg(insight_accuracy) FROM Actions WHERE led_to_insight="true" AND feature=f) AS insight_accuracy FROM Actions GROUP BY feature ORDER BY count(1) DESC

Platform	insights found by time spent	insights found by times used	insight quality by time spent	insight accuracy by time spent
form analytics	15.23%	55.88%	0.62%	0.68%
heatmaps	17.67%	44.76%	0.91%	1.10%
on-site polls	13.97%	29.23%	2.28%	2.75%

As we can see in table 5, quite logically, heatmaps provide insights the fastest. Form analytics, on the other hand, provided insights most often, 55% of the times they were used. When we add duration to the equation, on-site polls feature is a clear winner. This means, with least amount of time, using on-site polls, an analyst can find good quality and accurate insights. While form analytics provides good insights, analysing them takes significantly longer than heatmaps or polls.

If we combine the information from table 4 and table 5, the best combination is using the on-site polls feature in Hotjar as it provides the best insights the fastest. On the other hand, using Google Analytics to gain insights about form analytics takes the longest. Comparing the data types by platform (table 6), we can see the following trends: for heatmaps, Hotjar is the most accurate while the insights found from Mouseflow are the most actionable. For form analytics, Decibel Insights provides both high quality and accurate results. For on-site polls, again Hotjar seemed to work the best. If we would completely ignore the time spent metric, instead of heatmaps Google Analytics would be the most accurate solution for getting insights of where people clicked. Clicktale would work the best for form analytics and Decibel Insight for on-site polls. As seen in table 6, Google Analytics does not provide qualitative data (on-site polls).

Table 6. Platform comparison metrics by feature. Highest values marked with green. Source: Composed by the author.

Platform	insights found by time spent	insights found by times used	insight quality by time spent	insight accuracy by time spent
Heatmaps				
Decibel Insight	36.11%	68.42%	6.84%	6.20%
Reflective Data	14.58%	38.89%	3.27%	6.25%
Hotjar	29.41%	29.41%	10.59%	15.29%
Google Analytics	8.57%	52.94%	2.75%	4.02%
Mouseflow	31.82%	53.85%	14.29%	13.64%
Clicktale	11.11%	18.18%	11.11%	13.89%
Inspectlet	20.00%	40.00%	12.50%	13.75%
Form analytics				
Reflective Data	20.83%	58.82%	3.44%	3.91%
Inspectlet	14.43%	50.00%	3.53%	3.17%
Google Analytics	7.51%	50.00%	1.38%	1.42%
Clicktale	17.02%	57.14%	6.65%	8.51%
Mouseflow	27.27%	69.23%	9.09%	11.11%
Hotjar	18.75%	50.00%	9.38%	10.94%
Decibel Insight	28.57%	66.67%	15.08%	18.25%
On-site polls				
Clicktale	19.23%	33.33%	9.23%	12.31%
Reflective Data	11.11%	28.57%	9.03%	11.11%
Mouseflow	20.83%	41.67%	13.33%	15.83%
Hotjar	8.33%	10.00%	33.33%	25.00%
Inspectlet	12.50%	28.57%	18.75%	25.00%
Decibel Insight	11.11%	33.33%	22.22%	25.00%
Google Analytics	0.00%	0.00%	0.00%	0.00%

Based on the comments participants gave during the Think Aloud Protocol sessions, several common trends stood out. Many of them were also backed by the data of the actions they took. Most really liked the accuracy and flexibility of working with raw numbers but the extra amount of time is hard to find. Most of the user behaviour analysis tools do not provide the flexibility and custom segmentation options the analysts can find in tools like Google Analytics. Therefore a common practice is to use to tools with great visualisations, in most cases heatmaps and form analytics for finding bigger trends and problems of the website and then using tools like Google Analytics and on-site polls to better understand the issue. On the other hand, everyone agreed that general purpose digital analytics are just too slow for exploring new insights and this is the key reason why everyone is using at least one user behaviour analysis tool, the meaningful visualisations do make a difference.

Another thing users really liked is the simplicity of the visualisations in Hotjar, three of them also mentioned the good choice of colours. Two participants pointed out the limited data Hotjar provides in its form analytics section. They especially liked the advanced form analytics features in Reflective Data, pointing out the visual representation of problematic fields. About the tool called Mousflow, the participants liked its attention heatmap that visually represented the areas which get the most attention, and this is something other tools do not have every once in a while forces them to use Mouseflow. Four users mentioned that the “interactions dashboard” in Decibel Insight and its little graphs provide a lot of good insights that can later be further researched in other tools, this is also supported by the fact that 64% of the times participants used the heatmap feature in Decibel Insight they also found some sort of insight, their accuracy, on the other hand, was not significant.

For on-site polls, users tend to like tools which provide three types of data representations: table of responses, word cloud and word frequency bar charts. Three users mentioned they really liked the visual setup in Reflective Data and found while two found the one in Mouseflow not very pleasant to use. Two users also mentioned they would like to see a word cloud where they could easily exclude common words, and most tools already exclude words like “yes”, “and” etc. but common words may vary from business to business.

Five participants thought it is important that with every visualisation there is an option to download and share the visualisations. The more sharing options, the better, many tools had a “save as an image” option, some also had a “save as pdf” and only two tools currently offer raw data formats like CSV and JSON.

For heatmaps, three users pointed out that the visualisation should be configurable. Especially the opacity and colour scheme of the heatmap layer, one user also said that the point size in heatmaps should be customisable. At this point, only Reflective Data offers responsive heatmaps, making the visualisation a lot more flexible. The lack of this feature in other tools was mentioned twice. There was also a big difference in heatmaps rendering time, while some render the layer server-side, others do it in end-user's browser. The first might be a lot faster, but the level of flexibility is strongly limited. Two users said they would prefer to wait around 10-30 seconds more for that extra flexibility.

3.2 Discussion and Recommendations

Think Aloud Protocol sessions provided very good information about the mindsets and working habits of both digital analysts and online marketing specialists. The information collected gave a lot of new insights for reaching the goal of the thesis. The data collected in this research clearly suggests that data visualisation has a big impact on analysts work and helps to save a lot of valuable time for these busy specialists. As understanding user behaviour is one of the key actions in the process of improving the website and it can take a lot of time to find good insights, everything that helps is highly valued. Therefore, every company that is building a user behaviour analysis platform should put their focus on good visualisations.

When working on a new tool or improving the existing one, it is important to keep the end user in mind, and they should be involved in the process wherever possible. For example, semi-public alpha and beta testing rounds where users could easily submit their feedback. When looking at the platforms that were used by the participants of this research, it is easy to notice the ones that look rather outdated, their design seems to come from the mid 2000s, and the visualisations are robust and not very flexible. For example, the heatmaps are not taking mobile devices into account. When comparing the looks and the data that was collected in the Think Aloud Protocol sessions, we can clearly see these tools are performing worse in any given metric. At the same time, nice-looking, flexible and fast platforms also provided better insights faster.

Participants mentioned that at this point there is no tool that is good at both the visualisations and the raw numbers. This is the main reason why they have to jump between multiple platforms during the day, even when working on solving a small and very specific question. Everyone in the sample of this research agreed that Google Analytics is good at showing the numbers and very basic charts. They also agreed that most tools that specialise in the user behaviour analysis have good visualisations for form analytics, heatmaps and on-site polls but they lack the level of robustness that is often needed for further analysis and for proving what seemed like a promising insight in a visualisation. Therefore, this might be a good business opportunity for both types of platforms, to learn and implement what is best in the other one.

Some tools do not provide the option to see and download raw numbers because they are afraid that the customers will take the data and leave the platform. Others (like Google Analytics) do it so that users would have a motivation to upgrade to more expensive plans, where the feature is available. While, to some extent, it might work, there are two platforms that offer this feature in their basic plans and the users really noticed and liked it. If not having this feature is not a well-studied business decision, the author strongly recommends considering including this feature in one's platform and find another reason why the users should keep on using the software.

The participants of the Think Aloud Protocol really like heatmaps. The number one reason for that is that they provide rather accurate insights in a time that is not comparable with other solution on the market. The critical problems some of the tools have today are the lack of responsiveness (custom screen widths), limited customizability (opacity, colour scheme, point size) and the lack of data export feature. Users also mentioned they like the blue to red colour scheme much more than a range of shades of a single colour. The latter is also suggested by the concept of visual hierarchy.

One thing also stood out is that the participants liked innovative solutions that were only available in one or two platforms. Notably, the attention heatmap feature in Mouseflow and the "Interactions Dashboard" in Decibel Insight. Hence, as the field of visualising user behavioural data is rather new, there is no 100 percent right or wrong solution. The author would recommend companies to experiment with different solutions and gather user feedback to figure out what works best for their users. Same goes for the looks of specific visualisations, while they should at least somewhat match the company's overall branding, the user experience is what matters the most.

The Think Aloud Protocol proved to be a very successful method for this kind of research and will receive author's recommendations for anyone doing similar research or for companies that want to improve their user experience. Even a very little sample size of three to six participants can provide a sufficient amount of valuable insights. It is also a good idea to have the participants not only use the specific platform that is being studied but also its competitors. This way a researcher can get a broader overview of users' preferences and the areas where the competitors are doing a good or bad job. A method for storing the actions participants took in a SQL database worked well, too. It gave a good opportunity to run complex queries with rather a little effort and showed the results very fast. It also provided the ability to easily group, sort and filter the data based on any data-point that was stored in the database.

3.3 Limitations

Although the research was carefully prepared, there still existed some limitations and shortcomings. The first of them was the limited scope and resources of bachelor's thesis. A large institution could do a much more sophisticated research.

Secondly, this research focused on three main types of user behavioural data. The data that could be collected and later visualised is virtually limitless. Therefore, this research did not cover the whole spectrum of different visualisation solutions available for user behaviour analysis.

Thirdly, while the sample size of six provided a good amount of data, there is a chance that with some other six experts the results could have been totally different. The fact that only the experts with experience of using different tools were chosen should lower that risk by a lot, though.

Another aspect that in some way might have skewed the results is the fact that the participants in the Think Aloud Protocol sessions were all somehow related to the author of this research. This had some effect on the professional background of the participants. For example, there were no in-house enterprise level experts involved but rather people from agencies who are offering services to those enterprise level clients. In-house experts have known to have somewhat different working habits than people from the agencies.

As the topic is rather new, the amount of previous research available is very limited. There is a lot of research on data visualisation in general and measuring end-user experience on a website, but no research on measuring analysts' user experience in a digital analytics platform and how is it affected by different data visualisation solutions.

CONCLUSIONS

The goal of this exploratory research was to study different methods of visualising website user behavioural data in a digital analytics platform context and to answer the research question: how significant an impact does data visualisation have on the daily work of digital analysts and online marketers? The visualisation methods included in this research were those found in most popular user behaviour analysis tools (table 2). The main purpose of this paper was to study the relationship of data visualisation aesthetics and raw data by measuring the perceived user experience satisfaction level in a digital analytics platform.

To collect data, six Think Aloud Protocol sessions were conducted. All the sessions were transcribed, and then all the actions and other related information was stored in a SQL database. This allowed flexible queries and faster analysis. The sample included digital analytics and online marketing experts with at least two years of experience. As with most Think Aloud Protocols, the participants received very minimal guidelines and were free to use any tool they wished, as long as it had at least one of the three data types in place. Think Aloud Protocol was used for collecting both qualitative and quantitative data.

Key takeaways and recommendations from this research are: (1) wherever possible, provide both raw numbers and good visualisations, (2) provide an option for downloading and sharing both the visualisations and the numbers, (3) heatmaps are by far the fastest method for visualising mouse-related user actions, (4) form analytics data should be visualised in multiple different ways, especially good is the problematic fields report, (5) the visualisations should be configurable and flexible, (6) real users should be involved in the product development process, (7) experimenting with new visualisation solutions can uncover a solution users will love,

(8) Think Aloud Protocol is a great way of gaining insights for improving the user experience of a digital product, (9) learning what users like and dislike in competing platforms will provide a lot of insights for improvements. Based on the feedback from the participants, Decibel Insight seems to be the best in following the takeaways mentioned above: it shows raw numbers, allows downloading and sharing visualisations, has advanced heatmaps and form analytics features, many visualisations are configurable, they often ask for user feedback, and are currently experimenting with several new approaches.

As the platforms are constantly improving and new features are frequently added, to companies that are building similar software, the author would recommend occasionally keeping an eye on the competition. Many tools have open development roadmaps and public product launches. Also, conducting similar real user-driven experiments gives a lot of insights of what the users really think about these tools and features. Often companies are focusing on just one tool when doing user tests, most commonly their own tool. Including multiple tools, or give the user the freedom to pick the tools, provides a much broader overview and the ability to compare what is good and bad about each platform.

For future research on the topic of visualising user behavioural data in the context of digital analytics platform, the author would recommend continuing the Think Aloud Protocol method while also experimenting with other, more known research methods. When it comes to studying previous research related to this topic, there are a lot of papers about general data visualisations, general online user experience and the concept of visual hierarchy, all of them could provide valuable information for better understanding and further researching this topic.

REFERENCES

Agnieszka, B. . 2009. Informative or Misleading? Heatmaps Deconstructed. Human-Computer Interaction. New Trends. HCI 2009. Lecture Notes in Computer Science, vol 5610. Springer, Berlin, Heidelberg

Bitkulova, R. 2017. *Digital Marketing in Travel Industry.* Case: Hotel Landing Page Optimization. Saimaa University of Applied Sciences. 34

Bolton, H. 2017. Formisimo Academy: Getting Started. https://www.formisimo.com/blog/formisimo-academy-getting-started/

Bonggun, S., Lee, T., Choi, J. D. 2017. Lexicon Integrated CNN Models with Attention for Sentiment Analysis. Emory University. 8

Bouguet, S., D. 2017. How to Prepare for the General Data Protection Regulation (GDPR). (online) https://www.bryancave.com/en/thought-leadership/how-to-prepare-for-the-general-data-protection-regulation-gdpr.html

Brooks, A. 2017. 7 Form Analytics Tools to Help You Increase Form Submissions (online) https://www.ventureharbour.com/7-form-analytics-tools-to-help-you-increase-form-submissions/

Card, S. K., Mackinlay, Jock D., and Shneiderman. 1999. *Readings in Information visualisation: Using Vision to Think.* San Diego: Academic Press

Chaffey, D., Patron, M. 2012. From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice*, 14, 30-45.

Charters, E. 2003. The Use of Think-aloud Methods in Qualitative Research, 78

Clicktale Solutions Overview (online) https://www.clicktale.com/solutions/

Cui, W. Wu Y., Liu S. 2010. *Context preserving dynamic word cloud visualisation.* Visualisation Symposium (PacificVis), 2010 IEEE Pacific. March 2-5.

Darmanin, D. 2016. The Questions We Love to Ask Users (And You Should, Too!) (online) https://www.hotjar.com/blog/2016/03/05/questions-love-to-ask-users/

Djamasbi, S., Siegel, Tullis, T. *Visual Hierarchy and Viewing Behaviour: An Eye Tracking Study, in Human-Computer Interaction, Design and Development Approaches*. Lecture Notes in Computer Science, 2011, (6761:2011). 331-340.

Duchowski, A. 2003. *Eye tracking methodology: Theory and practice.* London: Springer.

Elmqvist, N. 2016 (online) https://sites.umiacs.umd.edu/elm/2016/01/04/joining-infovis-community/ (16.10.2017)

Esch, P., Esch, L. J. 2013. Justification of a Qualitative Methodology to Investigate the Emerging Concept: The Dimensions of Religion as Underpinning Constructs for Mass Media Social Marketing Campaigns. Southern Cross University, Queensland, Australia

Feldsted, CJ. 1950. *Design Fundamentals*. New York: Pittman Publishing Corporation.

Fei, S. 2009. Fundamental Differences in Analytical Tools: Exploratory, Custom, or Customizable. *Visual Business Intelligence Newsletter.* September/October 2009

Few, S. 2013. *The Encyclopedia of Human-Computer Interaction, 2nd Ed: Data visualisation for Human Perception.* Hershey: The Interaction Design Foundation

Forrest, W. Y., Leeuw, J. Takane, Y. 1980. *Quantifying Qualitative Data*. Similarity and Choice, eds. ED Lanterman & H. Feger. Bern: Hans Huber. 150-179

Genise, P. 2002. Usability Evaluation: Methods and Techniques: Version 2.0. University of Texas.

GIS Dictionary (online). ESRI. Retrieved. https://support.esri.com/en/other-resources/gis-dictionary/term/visual%20hierarchy (12.08.2014)

Google Analytics Product Overview (online) https://www.google.com/analytics. May 2018.

Google Analytics Usage Statistics (online) builtwith.com. March 2018.

Google Analytics. 2018. Custom dimensions & metrics: Include non-standard data in your reports (online) https://support.google.com/analytics/answer/2709828?hl=en

Granka, L. A., Joachims, T., Gay, G. 2004. Eye-Tracking Analysis of User Behaviour in WWW Search. Cornell University. 1

Healey, C.G. 1996. *Choosing Effective Colours for Data visualisation.* Visualisation

'96. Proceedings. November 1.

Honkela, T. 2016. Analytics of Qualitative Data Using Machine Learning Methods. Arcada Analytics Workshop.

Hotjar. 2017. How do Hotjar Heatmaps work? (online) https://help.hotjar.com/hc/en-us/sections/115002308328-Heatmaps

Hotjar. 2017. How do Hotjar Polls Work? (online) https://help.hotjar.com/hc/en-us/articles/115011772568-How-do-Hotjar-Polls-Work

Hotjar. 2018. Data Safety, Privacy & Security. (online) https://help.hotjar.com/hc/en-us/sections/115003180467-Privacy-Security-and-Operations

Hotjar. 2018. Tagging Recordings. (online) https://help.hotjar.com/hc/en-us/articles/115011819488-Tagging-Recordings

Inspectlet Product Features Tour (online) https://www.inspectlet.com/feature/session-recording

Jackson, I. 2008. *Gestalt—A Learning Theory for Graphic Design Education.* International Journal of Art and Design Education. Volume 27. Issue 1. 63-69.

Jaspers, M. W. M., Steen, T., Bos, C., Geenen, M. 2004. *The Think Aloud method: a guide to user interface design*. International Journal of Medical Informatics (2004) 73, 781—795

Jeon, M. H., Na, D. Y., Ahn, J. H., Hong, J. Y. 2008. User segmentation & UI optimization through mobile phone log analysis.

Johnstone, C. J., Bottsford-Miller N. A., Thompson S. J. 2006. Using the Think Aloud Method (Cognitive Labs) To Evaluate Test Design for Students with Disabilities and English Language Learners, 4-5

Johnson, K. E. 1992. Cognitive strategies and second language writers: A reevaluation of sentence combining. *Journal of Second Language Writing*, 1(1), 61-75.

Jääskeläinen, R. 2010. Handbook of Translation Studies: Think-aloud protocol. Amsterdam/Philadelphia: John Benjamins Publishing Company. 372

Kluyver, T., Kelley, B., Perez, F., Granger, B., Bussonnier, M., Frederic, J. Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C. 2016. Jupyter Notebooks—a publishing format for reproducible computational workflows. University of Southampton, UK.

Kucan, L., Beck, I. L. 1997. *Thinking Aloud and Reading Comprehension Research*. Inquiry, Instruction, and Social Interaction. Review of Educational Research, Vol 67, No. 3. 271-299

Laja, P. 2017. How To Use On-Site Surveys to Increase Conversions. (online) https://conversionxl.com/blog/on-site-surveys/

Lewis, C. H. 1982. Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report). IBM. RC-9265.

Liikkanen, L. 2017. The Data-Driven Design Era in Professional Web Design. Aalto University. 52

Liebling, D. J., Dumais, S. T. 2014. Gaze and Mouse Coordination in Everyday Work. Microsoft Research.

Mattox, J. R. 2013. L&D Needs Net Promoter Score. *Inside Learning Technologies & Skills,* July, 5. 135- 140.

Webster, M. 2018. Metadata. (online) https://www.merriam-webster.com/dictionary/metadata

Minkara, O. 2017. Customer Engagement Analytics: Take the Right Step, Every Time. Aberdeen Group. 7

Morse, J. M. 1994. *Designing funded qualitative research*. In Norman K. Denzin & Yvonna S. Lincoln (Eds.), Handbook of qualitative research (2nd ed.). Thousand Oaks, CA: Sage. 220-235

Mouseflow Platform Features Tour (online) https://mouseflow.com/tour/

Nielson, J. 1994. *Estimating the number of subjects needed for a thinking aloud Test*. International Journal of Human-Computer Studies, 41(3), 385–397.

Norman, D. 2013. *The design of everyday things.* Revised and expanded edition. New York: Basic Books.

Reflective Data Product Features Overview (online) https://reflectivedata.com/features/

Saunders, M., Lewis, P., Thornhill, A. 2009. Research methods for business students. Fifth edition. Lombarda: Rotolito Lombarda. 152

Stringfellow, A. 2018. Top Tools for Data Scientists: Analytics Tools, Data Visualization Tools, Database Tools, and More https://www.ngdata.com/top-tools-for-data-scientists/

Vaht, M. 2015. The Analysis and Design of a Privacy-Preserving Survey System. Master's Thesis. Faculty of Mathematics and Computer Science, University of Tartu. 14

Wilkinson, L., Friendly M. 2009. The History of the Cluster Heat Map. T*he American Statistician.* May, 63

Wroblewski, L. 2008. *Web Form Design: Filling in the Blanks.* New York: Rosenfeld Media, LLC

APPENDIXES

Appendix 1. List of Free and Open Source Software Used in This Research

1. MySQL library version 5.1.69
2. phpMyAdmin version 4.0.10.8
3. MySQL server version 10.0.29
4. Jupyter Notebook (former IPython Project)
5. Python version 3.6.5
6. Pandas version 0.22.0
7. Matplotlib version 2.2.2
8. NumPy version 1.14.2

Appendix 2. Rules and Guidelines for the Think Aloud Protocol

The following rules and guidelines were followed for every Think Aloud Protocol session.

- The participant is allowed to use any user behaviour focused digital analytics platform that has at least one of the following features/data types: form analytics, heatmaps, on-site polls
- The participants are encouraged to double-check their findings with more robust data-type. For example in Google Analytics or querying raw numbers from the database.
- The participants should use these tools just as they would in their normal working routine.
- Before the session, the participants were asked to evaluate every insight they find on the five-point scale based on two criteria: quality/importance of the

insight and the accuracy of insight.

- The observer should not interfere with the participant in any way during the session. The only allowed interaction is reminding them to speak, approximately after 30 seconds of silence.

Appendix 3. Choosing the Tools for Data Analysis

As mentioned in paragraph 2.4 a set of tools for data science (Stringfellow 2018) was being used for storing and analysing data. When choosing the right tools, the author took into account the structure of the data, research methodology and his personal experience. The tools included MySQL database, Python (a programming language), Matplotlib for plotting, NumPy for mathematical functions and Pandas for data structuring and manipulations. Most of the analysis was done in a popular Jupyter Notebook data analysis environment.

The reason for choosing a SQL database instead of NoSQL was mainly driven by the desire of keeping the data nicely structured and relational. SQL also provides faster queries and is, in general, being used more widely. The other tools were chosen mostly due to being a good fit for this kind of research as well as their wide usage in the industry of data science (Stringfellow, 2018). The fact the author had worked with all of these tools before made the decision even easier.

It is also important to note that using such tools was entirely optional and similar analysis could be done in a spreadsheet such as Microsoft Excel.