Evaluating the usability of data preparation tools for self-service business intelligence users

Jim Ekanem
4333802

November 29, 2022

Utrecht University

Master Thesis
Human-Computer Interaction

Evaluating the usability of data preparation tools for self-service business intelligence users

Jim Ekanem
4333802

1. Reviewer	Ioanna Lykourentzou Department of Beta Sciences Utrecht University
2. Reviewer	Judith Masthoff Department of Beta Sciences Utrecht University
Thesis Supervisor	Ioanna Lykourentzou
Day-to-Day Supervisor	Jorden van Foreest

November 29, 2022

Jim Ekanem

Evaluating the usability of data preparation tools for self-service business intelligence users

Human-Computer Interaction, November 29, 2022
Reviewers: Ioanna Lykourentzou and Judith Masthoff
Thesis Supervisor: Ioanna Lykourentzou
Day-to-Day Supervisor: Jorden van Foreest

Utrecht University
Master Thesis
Heidelberglaan 8
Postbus 80125, 3508 TC Utrecht

Abstract

Within the field of business intelligence, casual users depend on power users to create reports to make data-driven decisions. This is due to their low technical skill, which can be defined by the tools and technologies they use. This research sought to determine which usability problems casual users face in Trifacta Wrangler when executing three data structuring tasks that are required for report creation. Furthermore, the aim was to identify how technical skill influences the usability problems they face.

A usability test was conducted with 8 participants working in Sales, Marketing, and Client Services at a dutch marketing automation company. Participants were asked to fill in a survey inquiring about their technical skill and to fill in the System Usability Scale rating their interaction. As a result of categorizing identified usability problems according to the User Action Framework, it was found that most usability problems belonged to the planning and translation phase of the interaction cycle. Contrary to previous research, participants' Excel skill influenced their capability to plan interactions negatively. The System Usability Scale revealed that this might be related to learnability as a usability criterium. In line with previous research, one participant with prior SQL experience recovered from the most severe planning issue due to their knowledge of programming concepts.

It can be said that to improve the self-service level of casual users, understanding their planning of data structuring tasks is crucial. Further research is needed to verify these findings by identifying tools and technologies used by a larger sample of casual users and having them perform data structuring tasks in various market-leading data preparation tools.

Acknowledgement

Firstly, I would like to thank my thesis supervisor, Dr. Ioanna Lykourentzou, for your feedback, and support throughout writing this thesis. I am very grateful to have had you as my supervisor. Prof. dr. ir. Judith Masthoff, thank you for taking the time to be my second reader, as well as for your teachings in the course of this master's program.

I am grateful for my former colleagues at teamITG where I conducted my internship. Thank you for teaching me about email marketing and challenging me to expand my HCI perspective in a new domain. Thank you to Jorden van Foreest and Arjan van Hartesveldt for your mentorship.

Finally, I would like to give a big thank you to my family who has always supported me. Thank you to my friends Etienne, Domi, Jona, and Paulo, for being who you are. I feel deep gratitude for our friendship.

Contents

1 Introduction

Over the last decade "self service business intelligence has received considerable attention from both business communities and academia" [DA17]. BI compromises "all methods and tools for translating raw information into a meaningful, convenient form" [IKS19]. Data analytics is about analyzing this information to understand it and improve decision-making [CCS12]. In this context, self-service business intelligence (SSBI) could be used to empower business users with little technical skill - hereafter referred to as casual users - to perform custom analytics without needing support from more technically skilled people such as IT - hereafter referred to as power users[IW11]. In the context of this thesis, **self-service** refers to casual users independently being able to derive information from data sources.
According to Alpar and Schulz, there are different levels of self-service [AS]. The lowest level is "usage of information", and refers to the ability of casual users - to access information such as reports prepared for them by power users[AS]. The second lowest level is "creation of information" which implies casual users have data access and are able to create information from prepared data by analyzing it[AS]. The highest level of self-service is the independent "creation of information resources"[AS]. This implies that casual users must not rely on unified data access to various source systems being provided by power users.
Let us now consider the problems that lead organizations towards more SSBI investment and adoption. According to Daradkeh, there are several key business drivers [DA17]. In a rapidly changing business environment, the needs of casual users continuously change too. Thus, IT departments struggle with fulfilling related infrastructure requirements on time. According to Stodder et al., this is especially relevant to small businesses whereas, IT in larger organizations struggles with slow information access [Sto15]. Because there is more data being generated nowadays there is also the need for businesses to use more analytics, which requires more people to work with that data. This can reduce the effectiveness of data processing activities and the satisfaction of the people needing access.

1.1 Self-service from a people's perspective

Having looked at SSBI from an organizational perspective, let us now turn to the tasks and needs of people that work in organizations. As previously mentioned, the user pool is commonly differentiated between casual users and power users.

Casual users are managers or business people working in departments such as sales and make decisions based on data [PW; AS]. However, "analytics can only be as good as the data" that is being used [Sto16]. This is because data is stored in different systems, internal and external to the business, and must be prepared by power users before being available to casual users through their BI tools [ZSV]. This makes casual users information consumers [Eck08]. Any changes to the data must be requested by casual users and may lead to a time-consuming exchange process because casual users understand their own requirements better than power users [HHK18]. As a result, decisions can sometimes be made without casual users considering all existing data [Mic+20].
In this sense, casual users depend on power users because of their lack of technical skill, which plays a vital role in why there is a need for SSBI [Mic+20; CE17]. According to Eckerson, through SSBI, casual "create exactly the reports they want, when they want them"[Eck09].

Power users can be in roles such as data scientists, IT experts or business analysts [Eck11; PW; AS]. They make up 20% of the BI users in an organisation and use their technical skills to produce information such as reports for the 80% of casual users [Eck08]. This imbalance between casual users and power users makes it difficult for power users to meet the requests and results in them spending less time analysing data and completing their own work [AS; Eck11].
IT departments specifically face the challenge of offering an efficient technical infrastructure to support the exchange processes between casual and power users and as mentioned can also be information producers for casual users [AS]. Most of the time in data analytic projects is devoted to data preparation [HHK18; AT]. Traditionally data is integrated into BI tools by IT using ETL operations (extract, transform, load) because casual users and some more technical analysts do not possess the skills to find and integrate the data into BI tools on their own [Sav14; Kan+12]. As companies become more data-driven and the number of people requesting data grows, IT becomes an unscalable bottleneck burdened by unspecific data requests. [HHK18].

1.2 The role of software usability

With usability playing a crucial role in SSBI environments, let us now turn to how usable software can contribute to SSBI [IW11]. Self-service data preparation tools attempt to resolve the introduced issues by enabling casual users to access and transform the data they need themselves and thus increase their level of self-service. However, most tools require the user to have programming experience, to be an expert in the dataset domain, and to have prior knowledge and understanding of the datasets and the data preparation goal [HHK18; HN20]. As previously described, these skills are not held by one person in an organization, but they are shared by casual users and power users. Research suggests that **data preparation tools still demand too much expertise from their users** because tools fail to implement intelligent solutions for data transformation operations and therefore don't enable casual users to execute end-to-end data preparation themselves [CE17; HN20; Mic+20].

With the presented issues in mind, previous research concluded that **to make the Data Analytics process more efficient, data preparation tools must become more accessible and usable for casual users** [HN20; CE17; IW11; IW11]. Therefore, the research presented in this thesis aims to discover **how the usability of data preparation tools must be improved to increase casual users' level of self-service**.

1.3 Scientific and societal relevance of this thesis

From a scientific relevance perspective, the field of HCI can provide a unique perspective on the introduced issues. In this case, Usability Engineering as a subfield of HCI and research methods such as usability evaluations can yield results that inform future software designs to be more human-centered. Within HCI, several researchers have devised evaluation criteria specifically for BI tools. However, only a few studies have been done where these criteria were used for evaluation. Furthermore, the landscape of usability research on data preparation tools is even more scarce and their methodology can be viewed critically. Therefore, it would be interesting to conduct qualitative research with casual users that focuses on the human-centered design of data preparation tools.

Let us now move to the social relevance of this thesis. Researching self-service BI from an HCI perspective can contribute to enabling data access and improving analytic capabilities within organizations. In SSBI research, the term **data analysis democratization** has been used to describe casual users with different skill levels achieving high levels of self-service and, therefore, shifting from consuming information to carrying out analyses themselves [AS]. Even though the terminology is inconsistent in BI research, Data democratization can be defined as empowering employees to access and understand data [ZG18]. According to Lefebvre et al., the main impediments to achieving data democratization are a lack of data education among employees and limited access to data and analytics tools [LLF21]. In addition, they postulate that one of the main enablers of data democratization is a self-service data platform. Therefore, evaluating the use of self-service platforms from an HCI perspective can **contribute to the advancement of data democratization**.

1.4 Research questions

This research aims to evaluate the usability of a data preparation tool to identify how casual users' technical skill - relating to their capability to conduct data operations - influences the type of usability problems experienced when conducting data structuring tasks. With this in mind, the two research questions of the research are.

RQ 1: What **usability issues** do casual users face during data structuring tasks?
RQ 2: How do **technical skill** influence the performance of casual users on data structuring tasks?

The first research question emerged from the lack of research investigating the usability of data preparation tools and self-service business intelligence tools in general. The second research question emerged from knowledge gaps identified in related works, wherein no prior research has identified how technical skill influences the performance of data operations. The specific data operations identified in the related works are data structuring tasks that are crucial for enabling casual users' to independently create reports.

To answer these questions, a think-aloud usability test with 8 teamITG NL employees was conducted. TeamITG is a marketing technology company based in Utrecht. The participants were sampled based on their roles, which gives an estimation of the technical skill they possess [CE17]. Their task was to fill in a survey about data experience and conduct 3 data structuring tasks in the data preparation tool Trifacta Wrangler. Afterward, they filled in the System Usability Scale. The transcripts of participants' verbalizations were analyzed to filter out usability problems and categorized using the User Action Framework. The results are presented by investigating the relation between, on the one hand, the occurrence and categorization of usability problems and, on the other hand, the technical skills of the experiencing users. In the discussion section, the resulting insights are related to other researchers' findings. Furthermore, the scientific and social implications of the results in business intelligence and HCI are discussed. Lastly, the discussion and conclusion chapter provide an overview of the study's findings and present new research avenues to be studied in the field of HCI and BI.

2 Related Work

2.1 Big data processing for business intelligence and analytics (BIA)

In this section, we will look into different architectures used by organizations to process BI data and assist their decision-making process. Furthermore, we will look at the roles of the people that are involved and what challenges they face.

2.1.1 BIA architectures: From data to decisions

There are several architectures that depict the steps and stages of data processing for BIA decision-making within organizations. In this section two architectures, one from academic research by Passlick et al, and the other from a business report by Eckerson et al. will be reviewed [Eck11; PLB]. As can be seen in figure 2.2 and figure 2.1, both architectures depict systems, users, and how data flows among them. Passlick et al. focus on the actors and their skills, and they specifically model a stage in which analysis is performed [PLB]. In contrast, Eckerson et al. put an emphasis on explaining how different user groups interact with the components of the architecture [Eck11].

Let us first turn to the architecture model proposed by Passlick et al. [PLB]. According to the author, organizational data originates from several source systems that can be of two types. On one hand, there are internal source systems such as a Customer Relationship Management System (CRM), and on the other hand, there are external source systems such as Social Media. Source systems' data is then integrated into the storage and analysis infrastructure that contains various systems such as Data Warehouses and Hadoop clusters just to name a few. For this integration step Extract, Transform, Load (ETL) or ELT operations are necessary. The next step in Passlick et al's. architecture is the semantic layer that comprises tools that unify users' data access to the various storage systems of the previous layer.

The data is then accessed in the presentation and analysis layer by users with different levels of technical skill. According to Passlick et al. dashboards are used by so-called business users - who are the lowest technically skilled - to request reports or investigate KPIs [PLB]. These requests are fulfilled by medium-skilled business analysts or so-called power users. They work with Analytics-Portals that may enable ad hoc, self-service, or advanced analysis using the data that is for instance stored in data warehouses. Power users may also request new predictive and prescriptive analytics from the highly skilled Data Scientists that operate on raw data to search for new insights. It is apparent that all of the roles involved in this process possess different levels of technical skill that influence their ability to work with data. This will be investigated in a later section with the goal of identifying tasks and other characteristics that may distinguish the roles within data processing for BIA.

Let us now turn to the architecture model proposed by Eckerson et al. and compare it to Passlick et al's. architecture [Eck11]. This architecture is not as detailed containing less text and unnamed layers. However, there are similarities to Passlick et al's architecture regarding the underlying process that is illustrated. Firstly, data is also integrated from source systems into storage and analysis infrastructures using ETL operations. From there, it is accessed by Power users for ad-hoc queries, or loaded into BI tools to be used by less technical so-called casual users for accessing reports and dashboards. In contrast to the other model, both types of users are said to operate on two different types of architectures. Casual users operate on a top-down architecture preventing them from using data operations beyond their technical skill and enabling them to explore data through dashboards. Power users that are more technically skilled operate on a bottom-up architecture and can combine their own data with data from source systems. Eckerson postulates that having one architecture for both user types would present a struggle to either one of them due to their technical skill levels requiring different amounts of freedom.

After investigating the architectures, it is evident that they take different perspectives on how organizations generate value from data. **Passlick et al. illustrate how differently skilled users collaborate and access various data representations [PLB]. Eckerson et al. focus on surrounding users with the right technical environment to match their needs [Eck11].** Both architectures consider the people and their characteristics to be crucial to the success of generating value from data. In this sense, **the tasks that a user executes seem to be defined on one hand by their needs, and on the other hand restricted by their skill**. Further investigating the needs, tasks, and skills of the roles involved could provide insight into users' pain points to determine how systems can best support them.

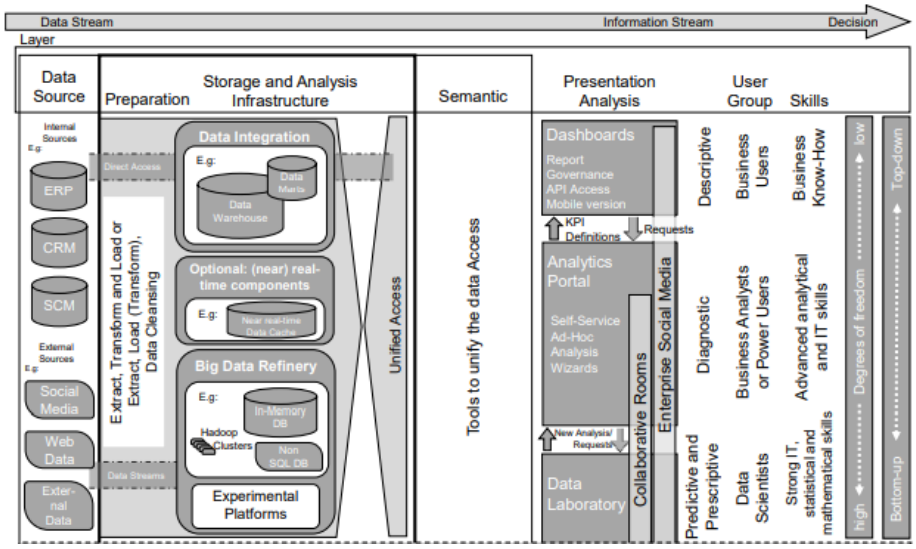


Fig. 2.1.: The BI architecture according to Passlick et al. Various user types access data through different interfaces. Their ability to do so is influenced by their level of technical skill [PLB]

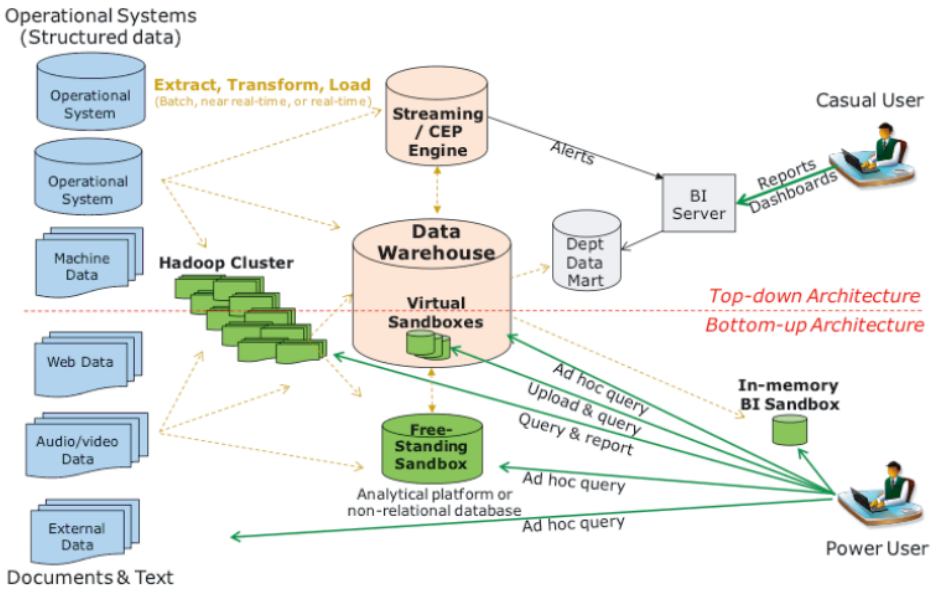


Fig. 2.2.: The BI architecture according to Eckerson et al. Casual users operate on a different architecture than power users when accessing data. The top-down architecture allows for less freedom which matches the casual users' low technical skill [Eck11]

2.1.2 Data processing stages

Having looked at the architecture within BIA, it is now necessary to better understand the challenges big data can present to organizations and the stages of data processing. In speaking of data challenges, the term big data is often used, which can be defined by several characteristics [GH15]. Volume is a characteristic that refers to data size, where large amounts of data may challenge a company's storage capacities. Variety refers to the structural heterogeneity of a dataset which may require new data management technologies and analytics to handle various types of data structures. Moving on from the variety of big data to velocity, which is another characteristic and refers to the rate at which big data is generated and the speed with which it must be used for BIA. On one hand, it may challenge the efficiency of communication in organizations when different roles collaborate to process data which could also be seen as necessary in the BIA architecture. On the other hand, velocity may slow down information access. The final two characteristics are veracity which refers to the unreliability of some data sources, and variability, which refers to data coming from different sources. This is important in BIA as data from different sources is often integrated and assembled in one system for a complete analysis [**20142014Report**]. This requires data to be prepared for unified access, as was mentioned in the previous section. The preparation process will be further detailed in the following section. This section has shown the characteristics of big data, of which each presents challenges to data processing. With big data being generated frequently nowadays, it is crucial that organizations can overcome the introduced challenges.

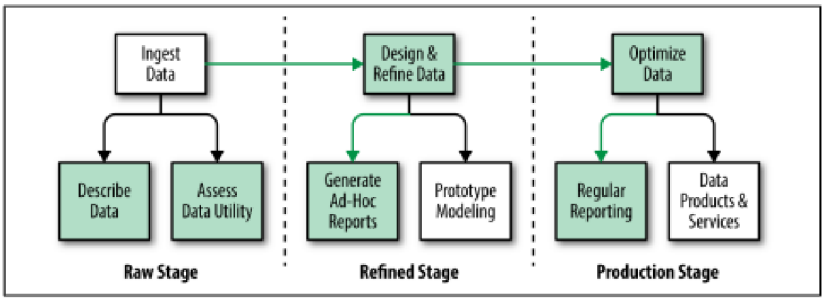


Fig. 2.3.: Stages of data processing and their contained actions according to Rattenburry et al. Highlighted are the Data Analysts' primary actions which show their ability to generate reports, a key task for independent decision-making [Rat+]

Let us now consider how data is processed for decision-making and which roles are involved. As shown in figure 2.3, according to Rattenburry et al., to gain value from big data, it moves through different stages that each contain three actions [Rat+]. In the raw data stage, the aim is to discover and understand the data. The actions are data ingestion, data description, and assessing the data's utility. The raw stage is followed by the refined stage, where data quality is enhanced, and unusable parts are removed. The related actions are designing & refining of data, generation of ad-hoc reports, and prototype modeling. Lastly, data enters the production data stage, where it is used for products and services, for instance, reporting and analytics. In this stage, the actions are data optimization, regular reporting, and data products & services.
Even though each action is a different building block to the bigger practice of processing data, the technical skills required from people to perform many of them are similar. This is because, according to Rattenburry et al., each action can require between one and five data operations in varying constellations [Rat+].

Let us, therefore, investigate the five data operations in detail. Three of them, namely structuring, enriching, and cleaning, are so-called data wrangling operations which are also mentioned in work by Hellerstein [HHK18]. Structuring changes the form or schema of data and thus can be related to dealing with the variety of big data. Enriching adds new values or combines datasets that can be related to the variability of big data. Cleaning fixes irregularities in a dataset which could be related to dealing with the veracity of big data. In addition to these data-wrangling operations, there are two types of profiling operations, namely individual value-based profiling and set-based profiling. These are both aimed at gaining a better understanding of the data. As previously mentioned, different compositions of data operations are necessary for different actions during data processing for BIA.

Therefore, it would be interesting to investigate which actions are conducted by which roles within an organization to identify which operations must be supported by computer systems. Rattenburry et al. describe this for several archetypes that they came up with, namely, Data Scientists, Data Analysts, Data Architects, and Data Engineers [Rat+]. The role of Data Scientists and Data Analysts are also mentioned in the discussed architectures by Eckerson et al. and Passlick et al. [PLB; Eck11]. In contrast, Data Architects and Data Engineers seem to not be involved in BIA and will therefore not be considered moving on.

Let us now turn to data analysts' tasks in each data stage. In the raw stage, the tasks are describing data and assessing data utility. In the refined stage, they are designing & refining data and generating ad-hoc reports. Finally, in the production stage, a data analyst's tasks are optimizing data and regular reporting, the latter of which both architectures mention as well. When comparing the data operations that are required for each of these tasks, it turns out that data structuring and both the profiling operations seem to be the most essential data operations for data analysts. However, at this point, the roles and tasks presented by Rattenburry et al. may give a one-sided perspective as they do not mention several roles that can be found in the discussed architectures [Rat+]. For instance, business users are excluded and based on Passlick et al. can be assumed to be even less technical than data analysts [PLB]. Therefore, it would be interesting to investigate how other researchers define the roles and tasks related to data processing within BIA. Especially regarding the data operations that each role can execute.

2.2 Power users and casual users: the actors within data processing

2.2.1 An overview of archetypes within BIA

Having looked at data processing within BIA from a technical side, let us now turn to the organizational side and, therefore, answer the question of who the people are that use BIA systems. As could be seen in the previous section and as other researchers have found, the literature makes non-coherent use of terms to name the actors within BIA. [JAC15]. However, most researchers use similar characteristics to describe each actor. Characteristics encountered are technical and business knowledge, tasks, and struggles. Therefore, it would be interesting to discuss the commonalities and differences in user archetypes characteristics' irrespective of the names and titles they are given. Based on this discussion, an archetype model will be derived to be used as a basis for this research. Regarding the following discussion, it is important to note that the field of pervasive BI may consider additional users of BI systems, such as front-line workers, suppliers, customers, and regulators [Wat09]. However, they will not be considered in the scope of this research.

It is assumed that BIA user archetypes were first described in a TDWI best practices report by Eckerson in 2008 and elaborated on by the same author in a 2011 report. [Eck08; Eck11]. According to Eckerson, users can be separated into power users who produce information and casual users who consume information[Eck08].
Power users have job titles like business analysts, analytical modelers, and data scientists. Their technical skills make them "savvy with software tools and familiar with applications and databases that are used to populate reports"[Eck08]. Daily tasks comprise crunching data to generate insights and make plans involving information access, analysis, and prediction. In addition, most ad hoc reports in an organization are created by power users.
Power users' struggles can depend on the architectural environment that surrounds them. In a top-down environment, they spend more time preparing data than analyzing it. This may result in data silos that are counterproductive to the organization's information consistency. Furthermore, within an organization, power users make up 20% of the BI users.
Subsequently, casual users make up the other 80% of BIA users. According to Eckerson, they "are not interested in learning about BI tools or databases", reflecting their technical skills [Eck08]. They are, however, interested in answering business questions.

To do so, casual users use predefined reports via the BI system's reporting function and dashboards that visualize data. They may also request the technically more skilled power users to create ad-hoc reports for them. This aligns with the findings from Rattenburry et al. and the architecture of Passlick et al [Rat+; PLB]. Furthermore, the dependence of casual users on power users illustrates their need to work in a top-down architecture with more assistance and less freedom.

Other researchers, Alpar and Schulz, and Phillips-Wren et al., categorized archetypes within BI into casual users and power users, too, while building on or amending Eckerson's model [AS; PW]. In their research, Alpar and Schulz investigated pervasive BI and enriched the model of power users and casual users based on their findings from related literature [AS]. Power users were explicitly labeled as a bottleneck during data processing which may be related to Eckerson's statement of them making up only 20% of BI users [Eck08]. Casual users' technical skill was described as SQL being too complex for most. Furthermore, casual users may not explore all the data available to them when making decisions or requests, which may impact the quality of their actions.

Phillips-Wren et al. conducted interviews with BIA practitioners and conducted a literature review, leading them to describe three archetypes [PW]. According to the authors, two archetypes equate to Eckerson's casual user and power user, which, therefore, will not be described again. The third archetype is the Data Scientist, who was previously categorized among power users. The reason for the distinction by Phillips-Wren et al. is that they depict data scientists as being more advanced in using data. Data Scientists possess strong math, statistics, and computer science skills and use them to create and deploy descriptive and prescriptive models. This relates to the prototyping activity in Rattenburry's refined stage [Rat+]. Furthermore, Data Scientists advise the organization in interpreting and visualizing data which relates more to the raw data processing stage.

So far, the archetypes illustrated were based on Eckerson's separation of actors within BI into casual users and power users. However, the specific requirements of analysts are only described by Phillips-Wren et al. [PW]. Therefore, in the following section, different archetypes found in the literature on data analysis will be discussed.

Several researchers and practitioners developed more varied archetypes representing users within BIA [CE17; Kan+12; Wat15]. In this section, they will be discussed in descending order of their characterized technical skill.

Kandel et al. interviewed 25 data analysts from various organizations [Kan+12]. As a result, they provided a more fine-grained view of data analysts by describing the tools they use and categorizing them into three archetypes: hackers, scripters, and application users that vary in terms of "programming proficiency, reliance on information technology (IT) staff, and task diversity, and vary less in terms of statistical proficiency" [Kan+12].
Hackers were the most skilled technically, knowing multiple programming languages and being able to chain together scripts that operate on different sources. Their tasks are transforming data and completing flexible workflows, which they achieve without help from IT. Kandel et al. did not mention any struggles [Kan+12].
Scripters' technical skills vary, with some being able to use SQL or write scripts. However, they cannot parse log files or scrape web data but they can perform advanced statistical methods. Their tasks involve manipulating data, such as aggregation and filtering, and using data for analysis that IT pulls from the data warehouse. Subsequently, they struggle with having to rely on IT to get the data they require.
Application users are the least technical and possess minimal to no programming skills. Their tasks include visualization, reporting, and also using data prepared by IT. Therefore, their struggles are the same as those of Scripters.

Convertino et al. conducted user research over a span of three years with business users that conduct data analysis tasks [CE17]. Furthermore, they investigated related work amongst others by Kandel et al. [Kan+12]. As a result, they distinguish between three archetypes, namely Data Scientist, Business Analyst, and Data Analyst. In addition, they described the tools that are used by the different archetypes.
Data Scientists resemble Kandel et al.'s. Hackers as they are the most technical users possessing programming and scripting skills. Their tasks are data preparation, meaning cleaning, filtering, and merging data sets that can be seen as a more elaborate description of Kandel et al.'s data transformation [Kan+12]. Furthermore, Data Scientists conduct statistical analysis, which, unlike data transformation, is also mentioned in Phillip et al.'s. Data Scientist archetype [PW]. However, in contrast to Hackers, Data Scientists use advanced tools and programming languages.
Moving on from the Data Scientist, Convertino et al. describe the Business Analyst that is less technical in comparison as this user is only able to conduct basic SQL operations. Most of their time is spent on data preparation tasks and the rest on analysis tasks.

They struggle with having to rely on IT, which equates to Kandel et al. Scripter's struggles [Kan+12]. Regarding tool use, Kandel et al. and Convertino et al. identified different tools that their second most technical archetype uses [Kan+12; CE17]. The least technically skilled archetype Covertino et al. describes is the Data Analyst who primarily uses spreadsheet programs [CE17]. Kandel et al. specify applications such as SAS and SPSS also excluding technologies demanding more technical skill such as R or Matlab[Kan+12]. They spend most of their time on data cleansing and reporting. An interesting concluding observation on tool use is that spreadsheet programs such as Excel are used by all archetypes. However, the more technical archetypes use various, more specialized tools. In this sense, according to Convertino et al., tool use is related to the degree of technical skill that one possesses [CE17].

Lastly, Watson's article on cloud computing based on conference discussions yielded five different archetypes of BI users [Wat15]. In contrast to the other researchers, Watson's descriptions are brief but include each archetype's business knowledge.
The most technical user possesses the least business knowledge, which aligns with the architecture model in figure 2.1. At the end of the spectrum and the most technically skilled were, therefore, Data Scientists. Their skills were similar to those of the Data Scientists described by Phillips-Wren et al., with additionally possessing solid skills in machine learning as well as some skills in computer science and statistics [PW]. Their main task is looking for questions or hypotheses worth investigating by exploring data, and thus, data preparation takes up most of their time. Therefore, task-wise, they are similar to the Data Scientists described by Convertino et al. and the Hackers described by Kandel et al. [CE17; Kan+12]. The main struggles mentioned are the technical skill required for data preparation and their need for more flexibility and less control in using data compared to Data Analysts.
Moving on to the following archetype, the Power Analyst conducts the tasks of a data scientist while possessing slightly worse technical skills but more business knowledge. In contrast, there are BI and Business Analysts who have a good understanding of the organization's data but varying technical knowledge. They deal with warehouse data to conduct analytical tasks, which they use dedicated tools for.
Next, the Power end-user has good business knowledge and sufficient technical skills to access, perform analyses, and create reports. They use excel to answer questions for casual end users that they deliver in the form of reports. This resonates with Logi Analytics' finding that "IT considers the use of spreadsheets to be the most important modeling for business users" [**20142014Report**]. However, in their survey, business users reported that in addition to consuming prepared reports, it is crucial that they can independently conduct analysis and produce reports [Log14].

The above-mentioned, Casual end user consumes reports, monitors dashboards, and explores data through them. They possess high business knowledge, and their characterization resembles Eckerson's casual user and all other researchers' least technical users. Overall the description of Watson's archetypes lacks specificity. Therefore, it can only be assumed that further archetypes apart from the Data Scientist resemble other researchers' findings. For this reason, Watson's archetypes will not be considered moving forward.

Having discussed the archetypes described by different researchers, it is apparent that all models share characteristics such as technical skills, tasks, and struggles. With this in mind, the models can be compared to derive a perspective that will be adopted for this research. However, the textual representation provided above may make comparing the models cognitively difficult which is why the similarities and differences will be visualized.

Study	Archetype	Raw Stage			Refined Stage			Production Stage			Information consumption
		Ingest Data	Descibe Data	Assess Data Utility	Design & Refine Data	Generate Ad-Hoc Reports	Prototype Modeling	Optimize Data	Regular Reporting	Data Products & Services	Report requesting & generating
Kandel et al	Application users										
	Scripters										
	Hackers										
Convertino et al	Business Analysts										
	Data Analysts										
	Data Scientists										
Phillips and Wren	Business User (casual user)										
	Business Analyst (Power user)										
	Data Scientist										
Eckerson; Alpar nd Schulz	Casual user										
	Power user										
Rattenburry et al.	Data Analysts										
	Data Scientist										

Technical skill level
Low technical skill
Medium technical skill
High technical skill

Fig. 2.4.: A comparison of the user archetypes within BIA summarizing the research discussed in the related works chapter. Additionally, each archetype's tasks within the data processing stages are illustrated

Figure 2.4 summarizes user archetypes from the related literature reviewed in this section with the colors signifying their relative level of technical skill. Furthermore, their tasks are signified from the list of data operations described by Rattenburry et al. [Rat+]. However, these are exclusively information-producing tasks. In considering the least technical users such as casual Users it is necessary to consider report requesting and dashboard drilling. Therefore, these were added to the visualization as information-consuming tasks. As can be seen, the users with the lowest technical skills mostly consume reports, make decisions, and sometimes conducts data analysis. These people who are referred to as casual users by Kandel et al., often need expert support to access and prepare their data [Kan+12]. This creates issues due to experts being a bottleneck in organizations. Therefore, more and more casual users will require support in the future as businesses become more data-driven. Most researchers model medium technically skilled users as capable of consumption and production. Therefore, in this research, we will adopt the model of Kandel et al. that separates casual (application) users, scripters, and hackers [Kan+12]. According to Siegel et al. non-technical consumers make up 65% of Analysis and BI professionals while active consumers, that can be seen as Scripters make up 25% [Sie+13]. Since they struggle the most with depending on experts it could be beneficial to investigate their relationship with IT and how it contributes to these struggles.

2.2.2 The tasks that lead to IT being a bottleneck

In the following section, it will be investigated how the roles involved in data processing and analysis for decision-making collaborate. In particular, the focus will lie on the communication and artifact exchange between application users, scripters, and IT which are not described by the architecture diagrams introduced earlier.
Many researchers have come to the conclusion that the least technical users are supported by the more technical users including IT for their analysis and data preparation [LVS; Kan+12]. But even more technical analysts receive support from IT for Data preparation.
According to Kandel et al., the IT team helps analysis in several ways [Kan+12]. Firstly, they keep data within a centralised platform which by ingesting it from source systems using SQL. Therefore, if an analyst - who can be a casual user or a power user - needs certain data, they will request it from IT. This, however, may lead to a long exchange of requests due to the amount of data that can be stored in different source systems [Mic+20; Eck08]. Therefore, the collaboration between casual users and analysts on one side, and IT on the other side is necessary for data preparation.
Furthermore, "the IT team is responsible for operationalizing recurring workflows" [Kan+12]. This means that casual and power users request IT to automate the consistent move and transformation of data that is constantly updated in source systems. In this case, collaboration helps more technical power users such as data scientists with writing scripts. Additionally, IT assists medium-skilled power users with finding and understanding data. Overall the described functions of IT serve users with varyious technical skill levels in different ways. IT's task of data preparation for casual users is the most important because make up over 60% of BI workers according to Siegel et al. and Eckerson et al. [Sie+13; Eck08].

Therefore, this research will focus on improving the self-service of data preparation for casual users. According to Rattenburry et al., the data operation that is part of most data processing tasks is the structuring of data [Rat+]. Furthermore, it is structuring that is necessary for ad-hoc and regular reporting. Thus, by improving the usability of data preparation tools for data structuring operations it is possible to reduce casual users' dependency on power users for data preparation and report generation.

2.3 Selecting a data wrangling tool

In this section, several data preparation tools will be introduced and discussed regarding their fit for this study. Based on this Trifacta Wrangler was selected and introduced as the subject of the usability evaluation.

2.3.1 A comparison of tools

There are many tools that enable data wrangling. Microsoft Excel is widely used amongst business users and can be used to analyze and visualize a dataset to recognize patterns and trends [CE17; Kan+12; RH21]. As related work has shown it is also widely used within BIA. However, analysis of large datasets with Excel may not be possible with over 1 Million rows [Kan+12]. This presents an issue due to the volume of big data which is why Excel may not be considered a tool for evaluation. However, it is important to inquire about participants' Excel experience as it may be related to their skill level.

Moving on from Excel there are other tools specifically intended for large and varied datasets. According to Rosett et al., many of them require programming skills. They may be designed to meet what organizations demand from the market. However, this demand does not directly reflect the need of the employees working in the organization. Most organizations seem to select BI tools that meet the needs of power users and which, therefore, are too complex for casual users [Eck12; Mic+20]. According to Watson, an explanation could be that the software selection committees are usually comprised of power users [Wat09]. As described in the archetypes, power users use programming languages such as R and Python for statistics, machine learning, and data wrangling. Therefore, it would be important to inquire about participants' experiences with them in order to detect a potential impact on their required support when wrangling data.

In light of the many tools that are too complex to meet the needs of casual users it is important to not select one of these in this study. Tools that are designed for casual users must be Graphical User Interface (GUI) based. Furthermore, tools that provide low-code or no-code functionality assist users that possess minimal programming skills [RH21]. Reports published by Forrester and by Gartner review the market of GUI-based data preparation tools presenting several established software products [Lit17; ZSV17]. These are for instance SAS Data Loader For Hadoop, Trifacta Wrangler Enterprise, Talend Open Studio, or Alteryx Analytics.

To the best of this author's knowledge, there has not been much research on the usability of the mentioned data preparation tools. A study was found rating the usability of Talend Open Studio [Ste+]. However, the methods used in the study relied on subjective opinions or impressions and are not reasoned from an HCI perspective.

The only tool that sufficient academic literature could be found on is Trifacta Wrangler. This is because it was developed from a cooperation between Stanford University and Berkley University before becoming a commercial venture in 2013 [Kan13; Sta]. Since then the software product has changed a lot in functionality and design while advancing to one of the market leaders [ZSV17]. Therefore, the current tool Trifacta Wrangler will be introduced based on information from the company's website and the above-mentioned independent reports []. Usability research about Wrangler will be discussed at the end of this section.

2.3.2 Case study: Trifacta Wrangler

This subsection will provide an introduction to Trifacta Wrangler and describe the functionality and user interface of the version that is used for this study. In the starters edition, users are offered Basic connectivity to cloud storage and cloud data warehouses. This enables the extraction from and loading into for instance cloud data warehouses. For this study, the free 90-day trial of the Professional edition (version 9.3.0) was used. Despite the multitude of possibilities, the study's only relevant feature is data wrangling. In this sense, Trifacta Wrangler supports the structuring, enriching, and cleaning of data. Common tasks such as aggregations, regular expressions, or joins are enhanced for non-technical users through now-code and low-code operations. The user interface also offers predictive features by previewing likely transformations as can be seen in figure 2.5. In addition to wrangling operations, the tool also offers data profiling to better understand the data by showing interactive visual representations of data patterns which can be identified as green bar diagrams in the interface.

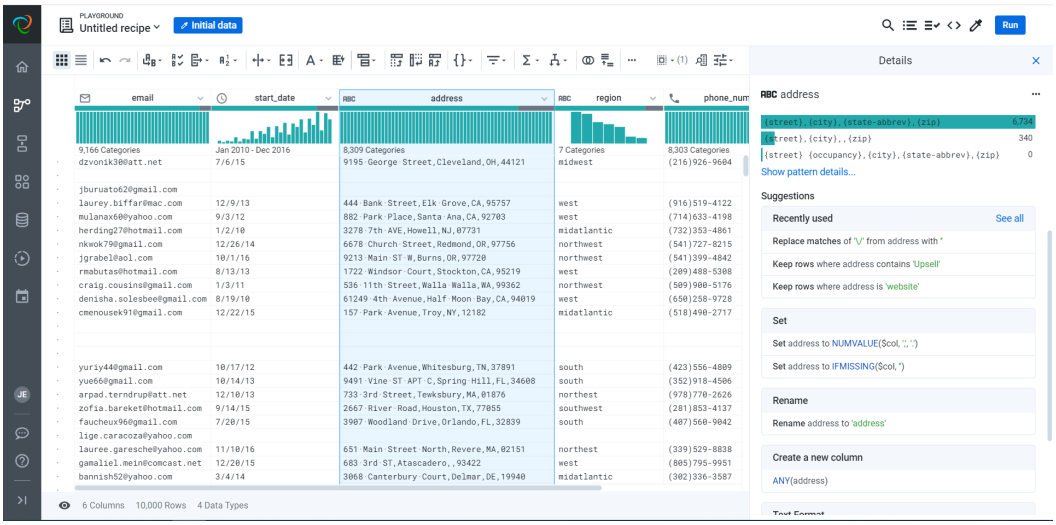


Fig. 2.5.: A screenshot from the data wrangling interface of Trifacta Wrangler []

2.4 Usability of data preparation tools

Having introduced Trifacta Wrangler as the tool for this study, let us now discuss the literature on data preparation tools' usability. As far as this author knows, usability evaluations of data preparation tools have only been conducted by Kandel et al as part of the development of Wrangler in 2011 [Kan+11]. Within their quantitative test, they had subjects perform three data-cleaning tasks with two different tools. The first tool was Excel and the second tool was the data preparation tool, Wrangler. They analyzed the influence of participants' self-reported Excel experience on their task performance. The result was that more Excel experience increased task performance. Participants' overall task execution was more efficient in Wrangler than in Excel.

To derive insights that inform this research, it is necessary to reflect on the above-described methodology critically. First, let us examine the tasks participants were instructed to perform and which data operation category the tasks are part of. As a reminder, the discussion of archetypes has found that casual users benefit from being able to do structuring operations on data. In Kandel et al.'s usability test, the tasks were text extraction, filling in missing values, and reshaping a table structure. While text extraction is a structuring operation, according to Rattenburry et al., reshaping the table structure is not mentioned by the authors and, therefore, uncategorized. However, filling in missing values is categorized as a cleaning operation, so Kandel et al. investigated the usability of multiple data operations. For this research's methodology, selecting three data operations from the structuring category is essential.
Moving on to the observed variables, Kandel et al. measured participants' task completion time. Thereby they addressed the effectiveness and efficiency of task achievement. In line with the definition of usability, it would be interesting to focus this research on collecting qualitative data about the user's satisfaction with the interaction. Doing so would help answer the question of why specific usability problems occur.

In addition to task performance, Kandel et al. measured excel experience, which makes sense when reflecting on casual users' tool use. Kandel et al. measured excel experience by self-reporting their estimation on a 10-point scale. Such a subjective measure may be prone to errors due to participants over- or underestimating themselves. However, to the best of the author's knowledge, there are no questionnaires to be found in related work that inquire about excel experience.

Therefore, when replicating this measurement, it is crucial to add questions that may inform the researcher about the participants' Excel experience in another manner. Furthermore, it would be interesting to measure the experience with other tools and technologies related to working with data too. Tools and technologies that Convertino et al. refer to as being used by data analysts include SQL, R, or Python.

2.5 Classifying usability problems

Having looked at usability methods used in research on data preparation tools, let us now turn to how usability problems are classified once they have been identified. Research on usability classification criteria and frameworks can provide useful insights. This section will present and discuss these insights concerning their usefulness for this research. Apart from general criteria, it is important to consider usability classification criteria developed for BI research.

In general usability literature, many criteria could be used to categorize usability issues and help understand why system designs are poor. Jakob Nielsen's 10 usability heuristics are frequently referenced criteria [Nieb]. Due to their popularity, they will not be detailed further at this point but instead taken into account when discussing BI-specific criteria at the end of this section.

Apart from this, researchers developed several frameworks to classify usability problems. On the one hand, the Usability Problem Taxonomy (UPT) enables classification at different levels of granularity [Kee+99]. On the other hand, the User Action Framework (UAF) is a support tool enabling the same with fewer levels of granularity. According to Andre et al., the User Action Framework is more reliable than the Usability Problem Taxonomy [And+00]. However, according to Khajouei et al., both frameworks are not sufficient for complete and accurate reporting as "it does not include a severity rating nor does it contain an assessment of the potential impact of usability flaws" [Kha+11]. Therefore, Khajouei et al. adopted the UAF's classification scheme and added elements of prioritization which can be seen in figure A.8 in the appendix. The latter prioritization first assesses the potential impact of usability problems on final task outcomes and secondly provides a validated severity rating rooted in Jakob Nielsen's work [Nie94].

Moving on from general usability literature, let us now turn to research findings specific to the context of BI.

In 2013, Jooste et al. came up with the following six criteria to evaluate BI applications: visibility, flexibility, learnability, application behavior, error control & help, and decision support [JVM13]. In proposing a new set of criteria, Smuts et al., 2015 combined the usability criteria by Biljon Mentz and Jooste et al. and added operability as a criterion [SSC15]. They also included the usability metrics satisfaction, efficiency, and effectiveness in their evaluation criteria. Therefore, when comparing these revised criteria of Smuts et al. with Nielsen's usability heuristics, seven heuristics can be seen as being contained in one of Smuts' criteria. Three of Nielsen's heuristics do not fit any of the descriptions: #2 Match between system and the real world, #3 user control and freedom, and #8 aesthetics and minimalist design. Based on the importance of Nielsen's heuristic in the literature, it can be argued to add them as criteria for evaluating BI applications.

Hameed Mousa et al. came up with six dimensions for measuring the usability of BI applications [HAT18]. However, they come with several limitations. On the one hand, several dimensions are particular to data processing and may not be generalizable to a more extensive scope of BI applications. On the other hand, one of the dimensions can be seen less as a criterion but rather as a feature. Unfortunately, the dimensions are not described in detail in the paper, which in any case prevents them from contributing to this discussion.

Looking at the introduced usability criteria for BI applications, it can be said that they are very similar to the 10 usability heuristics by Jakob Nielsen. This is probably due to the researchers' approach in deriving the BI-specific criteria from general usability research and combining them with results from their respective studies. Based on the discussion of usability problem classifications, an approach will be chosen in the following methods section.

2.6 Summary

In this section, the research gaps identified in the related literature will be presented once more and connected with each other.
Casual users possess business knowledge that they use to make decisions based on the organization's data. They possess little technical skill, which is reflected in their use of tools like Excel and other BI tools for which data is prepared by IT upon request. Casual users' technical skill restricts them to only being able to consume reports from these BI tools. Therefore, to reach a higher level of self-service and, thus, more independence from IT, they must be able to do regular reporting and generate ad-hoc reports. A key requirement for both types of reporting is conducting data structuring tasks in a data preparation tool.
Trifacta Wrangler is a market-leading data preparation tool originating from scientific research that can be used for a variety of data operations. Not much usability research could be found on BI tools, and no previous studies focused on issues casual users face during data structuring tasks. Therefore, it would be interesting to conduct a qualitative usability evaluation of Trifacta Wrangler to identify which problems casual users face and to classify the problems with one of the introduced usability criteria.

3 Methodology

3.1 Data collection to evaluate usability

In this section, the several factors that must be considered when selecting the methodology for a usability evaluation will be discussed.

3.1.1 Evaluation environment and sample characteristics

Usability evaluations can have different characteristics depending on the researcher's needs. They can be conducted in a controlled environment, making them replicable, or in the field, making the conditions more realistic [Pre+11; Dum02]. Furthermore, systems can be assessed by usability experts or by real users. The latter option assesses the actual use of the system while experts assess whether or not a system upholds accepted usability principles. Frequently cited principles for expert evaluation are Jakob Nielsen's 10 usability heuristics [Nieb]. Studies on heuristic evaluation show that it is better conducted by experts than novices, and several people conducting heuristic evaluations independently of each other is preferred over only one person conducting the evaluation [NMB90; Niea]. Due to this research being conducted by a single researcher, a heuristic evaluation may not be suitable and an evaluation based on real users is chosen. Furthermore, due to the difficulty of sampling business users that conduct data operations daily, the controlled environment is chosen over a field study.

Having decided to sample from a population of real users for the evaluation it is necessary to define the sample size. This is a frequently discussed topic in usability research as the number of users being tested impacts the number of usability problems that can be found in a study. Finding all usability problems of a system may be ideal however, trying to do so is inefficient [NL93]. Nielsen and Landauer recommend four to five users to discover 85% of the problems. Perfetti opposes them with her findings showing that testing with five users fell short of achieving the 85% problem discovery mark and recommends six users [UIE].

In contrast, Lewis recommends not relying on so-called magic numbers as they would only apply to specific conditions [Lew14]. Instead, one should use formulas based on the cumulative binomial probability to guide sample size estimation. However, these formulas require defining the probability of the occurrence of usability problems which in the context of this study can not be done accurately. In light of the discussed findings, this research will test with eight participants.

The data collection was supported by TeamITG Netherlands, whose employees were recruited as participants during an internship at the company. TeamITG is a marketing technology company based in the province of Utrecht. The casual users will be sampled from the marketing, sales, and data departments as their business and technical knowledge are expected to align most with the archetypes introduced in the related works chapter. They were contacted through an email that was forwarded by upper management. It informed the recipients of the study's context and its purpose. Furthermore, it provided a means to describe the ideal characteristics that participants should possess. On one hand, they were supposed to have little experience in working with data, for instance, in excel or similar data-related tools. On the other hand, they were not supposed to be very well-versed in SQL or other data-related programming languages. The goal of this constraint was to avoid sampling developers or other more technically skilled people. Following this email briefing, several colleagues that seemed to fit the description were approached by the researcher in person in an informal way to ask if they were willing to participate in the study. This personal approach was chosen to ensure that participants were self-motivated rather than instructed to participate. The people that agreed to participate were provided with a link to a date-picker tool where they could conveniently schedule their usability test.

3.1.2 Usability data collection process

Having discussed the evaluation's test environment and sampling characteristics, let us now turn to the evaluation procedure and its protocol. This involves several aspects. Firstly, a method must be chosen to collect qualitative data on usability issues that participants encounter during task execution. Subsequently, tasks must be defined. Furthermore, it is necessary to create a questionnaire to collect data on participants' characteristics and data-related skills. Finally, it is necessary to describe the responsibilities of the facilitator and the setup of the study.

Let us now turn to discuss the data collection method. According to Lewis, several methods can be used to identify usability problems by collecting data about users' effectiveness, efficiency, or satisfaction during system interaction [Lew14]. Interviews, usability tests, and standardized questionnaires, just to name a few. All methods can be used to create an inventory of usability problems, and there is no significant difference between the problems found in usability tests and expert reviews [MD08; Jor+96].
However, Jordan et al. postulate that only usability tests can measure task performance [Jor+96]. Concerning their reliability, a 2004 study by Molich et al. showed that independently conducted, identical usability tests, each found different usability problems, however, never the complete set of problems which is in line with Nielsen and Landauer's findings discussed when deciding upon the sample size [Mol+04; NL93]. Thus, the researchers recommended focusing on productivity and accurate usability classification schemes rather than maximizing the quantity of identified usability problems. Classification schemes have already been discussed in the related works section and will be decided upon later in the analysis section. Based on this paragraph's discussion, the usability test is chosen as the method to collect usability problems. Therefore, it must now be discussed how the usability test will be set up.

To identify usability problems, participants' problem-solving processes must be studied. According to Someren et al., there are several methods to do so [MW 94]. They can be characterized as observation methods, and verbalization methods. Let us first discuss observational methods such as recording eye movement, brain activity, or mouse clicks during system interaction. In the case of this experiment, recording eye movement or brain activity are not feasible due to resource limitations. Recording mouse clicks may be feasible however, doing so only provides information on what is happening on the computer system side of the observed human-computer interaction. However, from an HCI perspective, it is crucial to observe the problem-solving process from a human perspective.

Donald Norman's 7 stages of interaction model describes an interaction loop which is the same as the UAF's interaction cycle that is originally based on Donald Norman's theory of action [14; Nor87]. Within the 7 stages of action, the human's mental model of the world interacts with the computer's conceptual model through the interface. Any interaction may change the system state, which prompts the human to make sense of the change and perform subsequent interactions until the goal is achieved. Thus, the aim when collecting usability problems is to understand the mental model of the user and identify when the conceptual model does not match it. This means that recording click streams are not suitable for this research as they do not inform about the mental model.

Let us, therefore, turn to the verbalization methods where Someren et al. distinguish between unstructured and structured methods [MW 94]. Structured methods, such as asking a predefined set of questions to the user during the experiment, have several disadvantages. They require translating the subject's cognitive processes into a structure that may distort the information. Therefore, an unstructured method will be chosen where the information must be structured by the researcher, which will be discussed in detail in the following section.
There are several types of unstructured verbalization methods. During retrospection, the subject recalls cognitive processes that happened during the problem-solving process after finishing the tasks. This has the advantage that the subject can focus entirely on the task during execution. It is also reported that more problems were detected compared to verbalization during task execution [VDS03]. However, this could be influenced by a finding from Someren et al., according to which retrospective verbalization may lead to false memories being reported [MW 94].

False memory may also occur during introspection, where intermediate stops in between the problem-solving process are chosen by the subject to recall the cognitive process. The method of prompting is where the researcher interrupts the process to ask questions. Even though this may lead to a deeper understanding of the cognitive process, it makes measuring the efficiency of interaction difficult. In addition to these introduced verbalization methods, there is the think-aloud method, which requires subjects to concurrently verbalize their thoughts during a problem-solving process[Nie94]. On one hand, the benefit is that there is neither a delay in recall nor an interpretation of the cognitive process by the subject. On the other hand, verbalizing thoughts may lead to a slowing down of task performance which researchers' findings have been very contradicting on [VDS03; Jää10; VES86]. Ideally, a mixed methods approach would be chosen by applying a concurrent think-aloud that is extended using a retrospective interview.

However, such an approach would significantly lengthen the experiment's duration to exceed 1 hour. Considering the subjects' motivation to participate and their participation during work hours, this is not feasible. Therefore, the concurrent think-aloud will be chosen as the method to gather usability problems by recording the subject's verbalization of their cognitive processes.

Having decided on the concurrent think-aloud, it is necessary to define the tasks participants execute in Trifacta Wrangler. As explained in the related work section, tasks must be data structuring operations. Therefore, several structuring operations, according to Rattenburry et al., can be referenced [Rat+]. These are value extraction, filtering records and fields, and aggregations. As the authors provide example tasks for each operation, these can be adapted to fit the test dataset provided by default in Trifacta Wrangler. A task is completed if the ideal outcome is produced by the system.

3.1.3 Questionnaire about demographic, skill and usability

The previously described think-aloud usability test will yield qualitative data about the usability of Trifacta Wrangler. This will lead to the identification of implicit usability issues. However, the data lacks quantifiable information on how participants perceive the system's usability. Therefore, it could be argued to add a questionnaire that measures usability to the study. The system usability scale (SUS) can provide such insights [Bro96]. Therefore, participants will be asked to fill it in at the end of the usability test. An online survey will be created to collect demographic and skill data, present the tasks to the participants, and have them fill in the SUS. The tool for creating the survey will be Qulatrics, as it is licensed by Utrecht University [Qua].

Let us now turn to discuss the other data that is collected from participants. In the chapter of the related work, it was identified that several characteristics determine the technical skill of casual users. It is essential to take technical skills into account when interpreting the results of the usability test. In practice, questions about technical skills must be asked before participants take the usability test, as afterward, they may be influenced by their experience. The skills that will be inquired about are based on the research by Convertino et al., and Kandel et al. that was discussed in the related work chapter [Kan+12; CE17]. These are, therefore, Excel experience and experience with programming languages and statistical tools. Asking participants to report on their skills can be a biased measure which is why Convertino et al.s. findings will be used to inform additional questions [CE17].

Study	User Role	Tools
Kandel et al	Application users	Excel, dedicated analysis applications (e.g., SAS/JMP, SPSS, etc)
	Scripters	R, Matlab
	Hackers	Tableau, Excel, PowerPoint, D3, Raphael
Convertino et al	Business Analysts	Excel, dedicated analysis applications (e.g., SAS/JMP, SPSS, etc)
	Data Analysts	Excel, BI warehousing apps, MS Access, FoxPro, SQL client
	Data Scientists	Excel, Python, Ruby, Java, SQL client, R, Octave, Matlab

Technical skill level	
	Low technical skill
	Medium technical skill
	High technical skill

Fig. 3.1.: Technical skill can be determined by tool use. Casual users use Excel and analysis applications, while more technical roles can be identified by their programming languages and statistical tools. [Kan+12; CE17]

They postulate that a person's tool use can be equated with their technical skills. Therefore, it will be asked how frequently participants use the tools and technologies they subjectively rate their experience on. Figure 3.1 summarizes the findings by showing a skill continuum of users characterized by tool use.

In addition to the overall experience, participants may be familiar with one or more of the specific data preparation tasks from doing them in other systems. To identify influential factors, it is necessary to inquire about familiarity with aggregations, text extractions, and filtering operations.

3.1.4 Conducting the usability test

This section will describe how to prepare, moderate, and conduct the usability test. As preparation for usability tests, a moderator's checklist is recommended, which can be found in the Appendix in figure A.5 [Bar20; Ros22]. Participants were shown a 5-minute video about Trifacta Wrangler called "Getting started with Trifacta Wrangler" to give them an introduction into the system they are going to use [Tri]. Moving on to describing the technical setup, Trifacta Wrangler was run in a Chromium Browser on a Windows laptop that was provided by teamITG. Subjects were provided a mouse to use as an alternative to the trackpad.

To identify flaws in the test design, a pilot study was conducted. Therein it was revealed that the participant was unsure of the relevance of the 5-minute Trifacta Wrangler introductory video to the tasks they would to be conducted. As a result, the script was adapted to inform participants that the video provides a general overview of the system and their knowledge of it would not be tested. After viewing the video and beginning with the first task, it could be observed that the participant's unfamiliarity, combined with the testing environment, acted as a stressor for the participant. Therefore, each participant in the real test was given 3 minutes to familiarise themselves by freely discovering a "playground wrangling environment," which was set up to contain a different dataset than the actual tasks. This may be seen as training in which participants can already get to know the system. The downside would be that increased knowledge of the system could lead to participants recalling specific interaction patterns during task execution. This could be considered a limitation; however, in practice, business users receive extensive software training to which the short exploration can not be compared [TP09]. Therefore, it can be assumed that the test may still identify many usability issues.

After a few usability tests had been conducted, insights from that experience were used to adapt the facilitation. During the time participants spent exploring the interface, they were thus asked to practice the think-aloud method as a preparation for the tasks. Another finding from the pilot study was that specific phrases in the questionnaire were challenging to understand or could be misleading. For instance, the labeling of SQL, R, and Python, as tools even though these are tools and programming languages was corrected.

Regarding the actual study, a few issues could have impacted the results. One participant was used to interacting with a MacBook and, instead of using the mouse as a pointer, settled for the trackpad and described it as *unusual*. At the end of the session, another participant reported that the test situation caused him to feel nervous. If no one had watched the participant perform the tasks, he would have tried solving them for longer and exploring more creative solutions. This aspect of the study setup will be reflected upon in the discussion. Furthermore, the screen recording of another participant was lost at the beginning of the analysis, which is why their session was analyzed solely based on the audio recordings.

Apart from the moderator's checklist, the guidance of a participant through a session also involved ad-hoc communication. Initially, participants were supposed to think out loud without interruptions by the moderator while trying to solve the tasks. However, occasionally participants had to be prompted to verbalize their thoughts when they forgot to do so consistently. Participants could also ask questions if they did not understand the task. However, questions regarding the system's function or how to complete a task were not answered. The 8th interview could not be conducted in person due to illness and was conducted online using Microsoft teams. The meeting software allows for screen control and recording and automatically generates transcripts. A downside of the online usability test is that occasionally the internet connection negatively impacts the software's responsiveness. However, the participant was made aware of this before beginning the tasks. Therefore, it can be assumed that the connection problems did not bias their impression of the system's usability.

3.2 Analysis

3.2.1 Classifying usability problems

This subsection will discuss the analysis of the usability test data. Based on the discussion in the related works the User Action Framework will be used to classify usability problems. This includes the categorization, severity, and potential outcomes of usability problems, as well as the transcript coding. Before discussing the above, the necessary preparations are described.
Firstly, the participant's think-aloud recordings must be transcribed. This will be done with OtterAI and Microsoft teams which use speech recognition technologies to generate transcripts automatically [Ott]. The result may still contain errors that will be proofread and corrected if necessary. To analyze participant interaction utilizing observation, the screen recordings will be matched with the voice recordings to produce a complete video of the system interaction. This video serves as the basis for the identification and description of usability problems. A sequence can be identified as a usability problem if the participant deviates from one of the optimum procedures for problem-solving [VDS03]. These optimum procedures were identified in Trifacta Wrangler before the experiment's conduction. The sum of problem descriptions makes up an inventory of usability problems that are numbered and tagged according to the participant's anonymous identifier and the task during which they occurred. Said usability problem descriptions can then be classified into the User Action Framework. When classifying each usability problem in the relevant area of the UAF, as many classification nodes as possible must be labeled to provide a fine-grained classification. The classification tree can be seen in figure 3.2. As preparation, different classification scenarios will be practiced on the UAF training modules website, [Bru]. For the usability problem classification, another HCI master's student was recruited and trained. However, in the two available hours, it was not possible to complete the classification tasks. The second coder managed to code half of the problems and stimulated a discussion around the UAF framework that resulted in a reclassification of more usability problems.

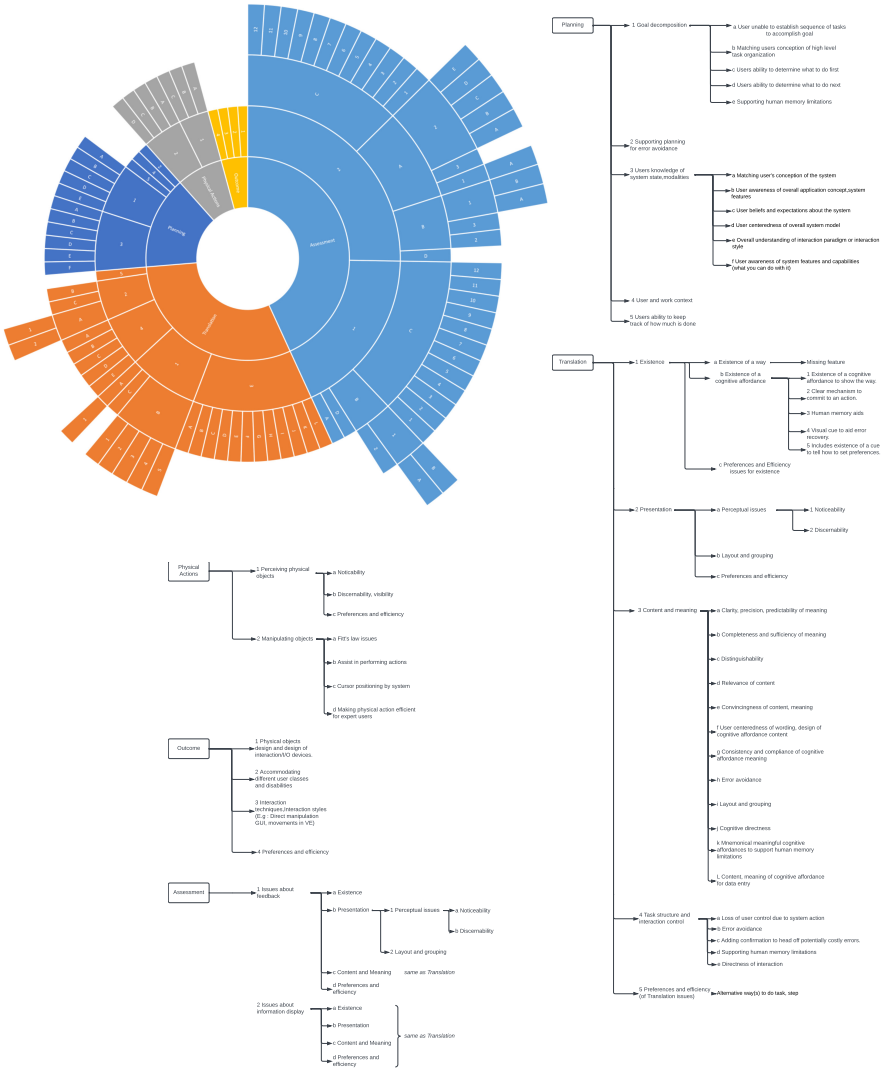


Fig. 3.2.: The User Action Framework phases including the subcategories and their specific codes next to a sunburst diagram of the classification tree. For improved readability the classification tree can be found in the Appendix figure A.10 and figure A.15

Usability problem classification is followed by rating its severity, which ranges from 0 to 4 and has been adopted by the UAF researchers based on Jakob Nielsen's literature[MN93]:
"0 = this is not a usability problem at all
1 = cosmetic problem only - need not be fixed unless extra time is available on a project
2 = minor usability problem - fixing this should be given low priority
3 = major usability problem - important to fix, so it should be given high priority
4 = usability catastrophe - imperative to fix this before the product can be released"[MN93]
However, the description that Nielsen gives on determining the severity is not accurate enough to do so reliably outside of an actual project's context. Therefore, an alternative rating of Nielsen will be used, which can be seen in figure 3.3. The rating is made up of **how many users experience a problem and what its impact is**. If the number of participants that experience the problem surpasses more than half of the participants, which is four, then it will be tagged as being experienced by *many*. The problem's impact will be determined by whether it impacts task completion or not. Therefore, the impact is *large* if it is the reason for a user to fail the task.

		Proportion of users experiencing the problem	
		Few	*Many*
Impact of problem on the users who experience it	*Small*	Low severity	Medium severity
	Large	Medium severity	High severity

Fig. 3.3.: The "table to estimate the severity of usability problems based on the frequency with which the problem is encountered by users and the impact of the problems on those users who encounter it" [Nie94].

Having looked at the severity, let us turn to the specification of a usability problem's potential outcome. This undertaking can be deemed unfeasible in this research because the usability test's tasks are not tied to an overarching project. Therefore, the potential outcome could only be assumed but not accurately estimated. As a result of this reflection, the final prioritization step, as defined by Khajouei et al., will not be considered [Kha+11].
The above-described analysis steps will yield an inventory of classified usability problems. To generate insights, the usability problem classifications will be grouped per participant and task. Tasks are classified as either completed or not. This provides an overview of which types of problems occurred most frequently in each task, in general, and what their severity is.

3.2.2 Coding the transcript

According to Fan et al., think-aloud protocols can be deductively coded with specific codes [Fan+19]. Sequences, where participants read text from the interface or task instructions, are coded as *reading*. Descriptions of participant's actions are coded as *procedure*. Remarks about the device or the participant themselves are coded as *observations*. Finally, there are explanations of motivation for participants' behavior coded as *explanations*. Fan et al. postulate that usability problems are most likely found in *observations*; however, they define no further steps [Fan+19]. Therefore, to gain a qualitative insight into the usability problems, open coding will be applied within the *observation* category. This is done by identifying verbal indicators of problems [VDS03]. In practice, verbalizations such as *I am lost* or *This was unexpected* were coded. In this case, the equivalent categories were named *noPlan* or *unexpected*. The verbalizations will then be used to verify the usability problem categorizations and to provide qualitative context to the results in the form of quotes.

3.2.3 Technical skill's impact on usability and SUS

To identify the effect that different levels of technicality have on usability, it is necessary to combine the survey results and the identified usability problems. It will be investigated if different technological experiences and frequencies of tool use impact task completion as well as the type of usability problems that occur and their severity.
Moving on from the usability problems derived from the test, the SUS questionnaire provides a subjective view of Trifacta's usability. This insight will complement the usability problem categorisations in determining if participants with different levels of technical skill perceive the system's usability differently. Therefore, it is necessary to identify correlations between the SUS score and the participant's technical ability. Furthermore, it is essential to identify connections between the SUS score and the occurrence and severity of usability problems to possibly inform future research on the SUS.

4 Results

4.1 Participant characteristics

Eight people participated in the experiments, four between the age of 35 and 44 and two between the age of 25 and 34. The two others were between 18 and 24 years old and 45 and 54 years old. The gender of six participants was male, while two were female.

Most participants reported being at least moderately familiar with filtering operations. In contrast, few participants reported being familiar with extractions and aggregations.

All participants reported having prior experience in Excel, with all but one using it weekly or daily. Between Excel experience and frequency of use, no significant differences could be found among the participants. Moving on to SQL, one participant reported having basic SQL experience, and the others were inexperienced. None of the participants currently used SQL.

Participants had no experience with R, Octave, and Matlab and had never used them. The same goes for Python, Ruby, and Java. In addition to the already-named technologies, four participants reported using PowerBI daily or weekly. Three of them worked in the Sales department, with the other participant working in the Client Services department. PowerBI is a business intelligence tool to connect and visualize data to gain insights [Mic]. Therefore, the influence participants' PowerBI experience has on their performance will be analyzed, too.

In conclusion, it can be said that casual users are not very familiar with data structuring operations. They use Excel frequently and have intermediate experience with it. In contrast, advanced programming and statistical languages are unknown to casual users, whereas SQL experience can be considered an anomaly among them.

4.2 Analysis encodings

Let us now turn to the encoding of the results. As shown in figure 4.1, several codes were produced during the analysis. The codes' usefulness varies because not all of them can be used to verify usability problems. For instance, *observationDescription latency*, *languageBarrier*, and *inferenceFromDescription* either do not reference problems or relate to issues with the experiment setup. The remaining codes were used to gain insights into usability problems. In practice, the code could hint at an interaction phase where the problem occurred. An example of this is the code *unclearMeaning* which can relate to the subcategories Content and Meaning of either the Translation Phase or the Assessment phase. In contrast, the code *unexpected* simply informed about a usability problem but did not hint at a specific phase.

Code	Description	References
inferenceFromDescription	Participants made an inference about the procedure from what they observed	25
observationDescription	Particpants described what they observed on the system side or on the screen	25
unexpected	The observations participants made in the system were not what they expected	23
unsureHowToDo	Participants were unsure how to execute what they were planning	18
negativeRemarks	Participants made negative remarks about the system and its interface	15
error	Participants encountered an error	10
selfBlame	Particpants blamed themselves after making an observation	10
noPlan	Participants did not know what to do next	9
unclearExpectation	It was unclear to participants what to expect prior to an interaction sequence	8
unclearMeaning	It was unclear to participants what the meaning of something they read or observed in the system was	6
languageBarrier	Participants had difficulties with the English language	2
latency	Participants remarkted connection issues during the remote usability test	1

Fig. 4.1.: The open codes from the observation category and their frequency of reference produced during the coding process in Nvivo.

4.3 RQ1: Most usability problems identified are about Planning and Translation interactions

Overall task performance varied among each of the three tasks. Six out of eight participants completed the extraction task by reaching the ideal outcome through one of the ways illustrated in the figure A.9 in the appendix. The filtering task was completed by seven out of eight participants. In contrast, the aggregation task was only completed by two out of eight participants. These two participants were also the only participants that managed to complete all three tasks. The difficulties participants experienced during the aggregation tasks can be observed in the transcript's codes. Participants blamed themselves mostly during the aggregation task, with verbalizations such as:
'This is basic, why am I not able to solve this?... I'm doubting myself right now.'
The code lost was mainly related to the aggregation task:
'I don't really know what to do differently' and *'I don't know what I'm doing at the moment.'*

The completion rate differs amongst tasks; however, the number of usability problems experienced was similar when comparing all completed and failed tasks with each other.
During the usability tests, 68 usability problems occurred. When looking at the usability problem categories in figure 4.2, it can be seen that most of the problems are categorized in the planning and translation categories, with two problems being in the assessment category. None of the usability problems could be attributed to either of the two categories outcome and physical action. The occurrence of problems from each category is almost evenly distributed amongst the three data structuring tasks. When investigating the problems that the two participants who completed all tasks experienced, it can be seen that amongst them, only one planning problem and one assessment problem were encountered. The other 11 are in the translation category. It seems as if translation problems have a lesser impact on task completion. Therefore, let us now turn to the usability problem severity to gain further insights into which problems impacted Trifacta's usability the most.

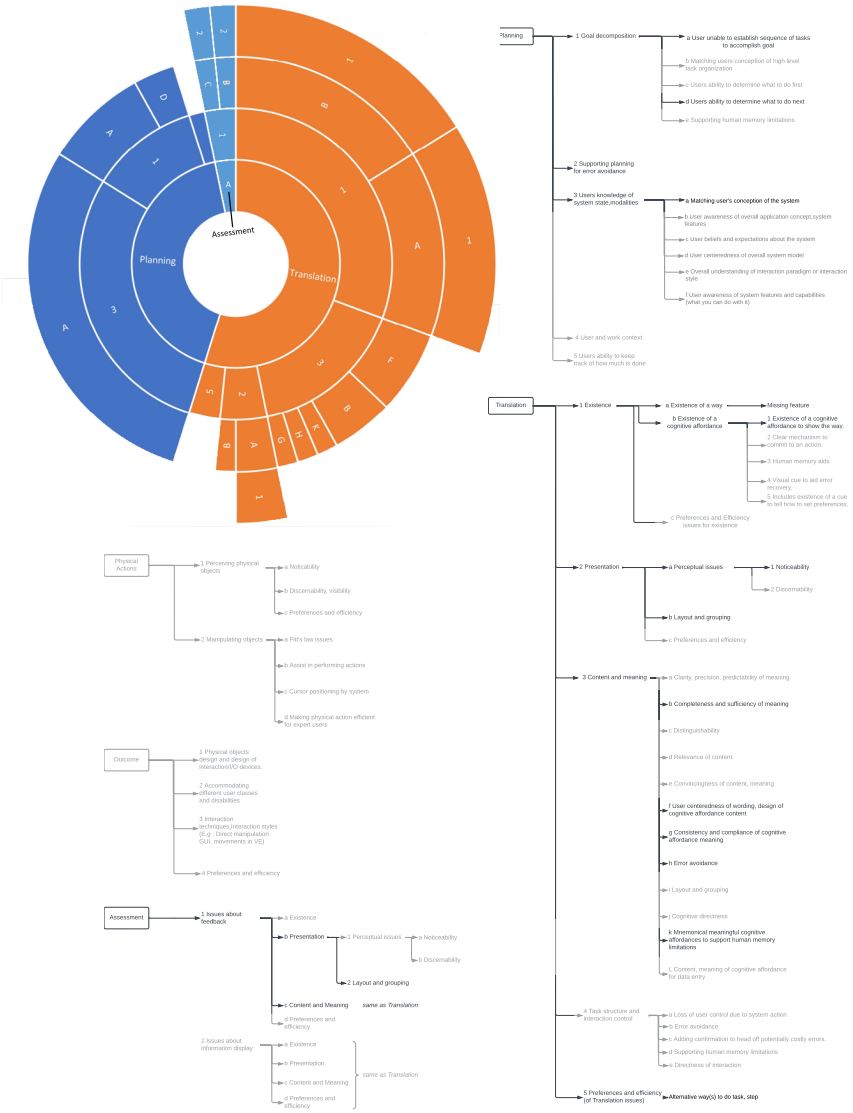
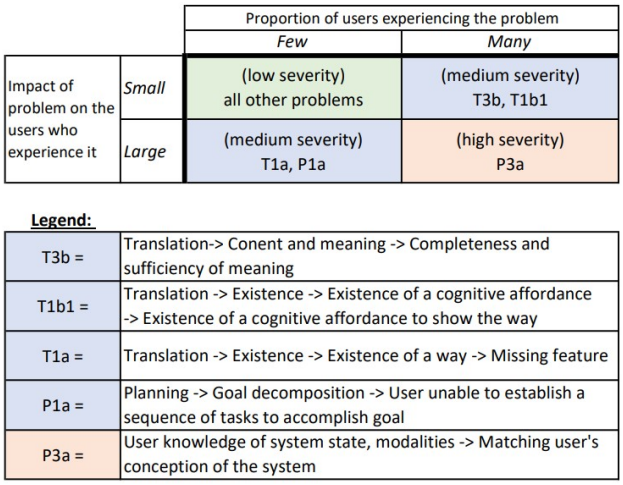


Fig. 4.2.: The results showing the User Action Framework phases of the identified usability problems in Trifacta Wrangler. Most problems are in the categories planning and translation and two are in the category assessment. The text highlights only those subcategories and codes that are visible in the sunburst diagram.

Figure 4.3, shows the usability problems with medium and high severity along with their codes in the UAF's classification tree. The axis' of the table indicates the proportion of users experiencing the problem and whether the problem led to task failure more than half the time. The most severe usability problem (P3a) is in the planning category. The problem relates to matching users' conception of the system, which was the most experience issue among participants and led to task failure most frequently. In all three tasks, the user could not formulate regular expression code according to the syntax required by the program. While the aggregation task demanded a more complicated formula, the extraction and translation tasks required inputting a single string between quotes to specify the value that should be extracted or filtered. Participants were unaware that strings had to be put between quotes as part of the syntactical rules. This is related to the medium severe translation usability problem (T3b) of the incomplete feedback message and its unclear meaning to casual users. Many participants mentioned this during the think-aloud session:
'I put in the slash as the delimiter, and then it gave an error message, which I didn't really fully understand.'
'I'm probably missing something there, but I'm not sure what it is.'
'And I'm getting the same error I got in the previous task saying column upsell does not exist in the schema. No, it doesn't.'



		Proportion of users experiencing the problem	
		Few	*Many*
Impact of problem on the users who experience it	*Small*	(low severity) all other problems	(medium severity) T3b, T1b1
	Large	(medium severity) T1a, P1a	(high severity) P3a

Legend:

T3b =	Translation-> Conent and meaning -> Completeness and sufficiency of meaning
T1b1 =	Translation -> Existence -> Existence of a cognitive affordance -> Existence of a cognitive affordance to show the way
T1a =	Translation -> Existence -> Existence of a way -> Missing feature
P1a =	Planning -> Goal decomposition -> User unable to establish a sequence of tasks to accomplish goal
P3a =	User knowledge of system state, modalities -> Matching user's conception of the system

Fig. 4.3.: The headlines of the table's row and column indicate the severity classification of the usability problem codes. In the legend, the full classification can be found according to the UAF. It can be seen that the most severe problem belongs to the Planning category and the majority of medium-severe problems belong to the Translation category. All other problems have low severity.

Moving on to the three medium-severe problems in the translation category (T3b, T1b1, T1a). They relate to missing features, missing affordances to show the way, and non-user-centered wording. The latter was expressed by participants when trying to decide between the call of actions 'day of the week' and 'day of the month':
'Don't know what day of month or day of the week is.'
'I was wondering which one is the best to choose.'
'I don't understand the difference between day of the week or day of the month, but it seems logical to do day of the month.'
The other medium-severe problem is in the planning category and refers to users' inability to establish a sequence of tasks which caused two participants to fail the aggregation task.

The think-aloud results are also confirmed by the System Usability Scale. It produces an average rating of 49 for Trifacta Wrangler. This is considered to be a poor rating and indicates that the system's usability must be improved to be usable for casual users. Each participant's SUS score rated the system's usability very differently, as seen in the appendix figure A.12. Investigating the poor SUS ratings showed similar answers to questions 3 and 10: Participants with low SUS scores **did not find the system easy to use** and, on the other hand, found that **they needed to learn a lot of things before they could get going with the system**. Apart from this, no connection between the SUS score and the data collected in the questionnaire could be found.

4.4 RQ2: Technological skill influences the occurrence of usability problems

This section will describe the influence of casual users' technical skills on their performance on data structuring tasks. Technical skill is characterized as participants' experience with and use of tools and technologies inquired about in the survey, or that were given as open answers.

4.4.1 Users with less technical skills are more likely to incur more severe planning issues

As shown in figure 3.1 in the method chapter, Excel was used to indicate the lowest degree of technical skill. It can be seen that participants with more self-reported excel experience completed more tasks. However, as there are no significant differences in Excel experience and frequency of use amongst the participants, no profound conclusion can be drawn about the impact of Excel experience on data structuring tasks. Looking at the transcript, however, it can be seen in figure 4.4 that Excel has been mentioned by multiple participants with varying sentiments. At the time of one participant's failure to complete the aggregation task, they frustratingly claimed to be able to conduct it in 30 seconds in Excel. Another participant exclaimed:

'Mm, it's hard to find the option that I think I need. To filter all the other options out, for example, in Excel.'

The quotes indicate the negative influence that Excel knowledge can have on the performance of aggregation tasks in Trifacta Wrangler. On the other hand, participants also reported that the user interface looked similar to Excel because of its menu bar that shows all functions and the large data table. One participant uttered: *'It's difficult for me because I never worked with a system like this, but yeah, it's a little bit Excel-ish. [...] the look and feel.'*

In light of these insights, Excel experience may influence the user experience, but not the usability of Trifacta Wrangler.

Fig. 4.4.: Several participants produced quotes related to their prior experience with Excel. The common sentiment is that the participants conducted their system interaction with Excel in mind and therefore encountered problems with Trifacta Wrangler.

4.4.2 Users with more technical skill in data operations experience fewer Planning issues and more Translation issues

In contrast to Excel, SQL experience could influence task performance and usability problems. The participant with SQL experience completed all tasks. Furthermore, as shown in figure 4.5 the transcript suggests that knowledge of SQL enabled said participant to recover from the highly severe planning problem. In that case, entering code according to the system's programming syntax was a less severe usability problem for said participant because, despite it, they completed the task. The quote in figure 4.5 provides the participant's explanation as to how SQL experience helped them.

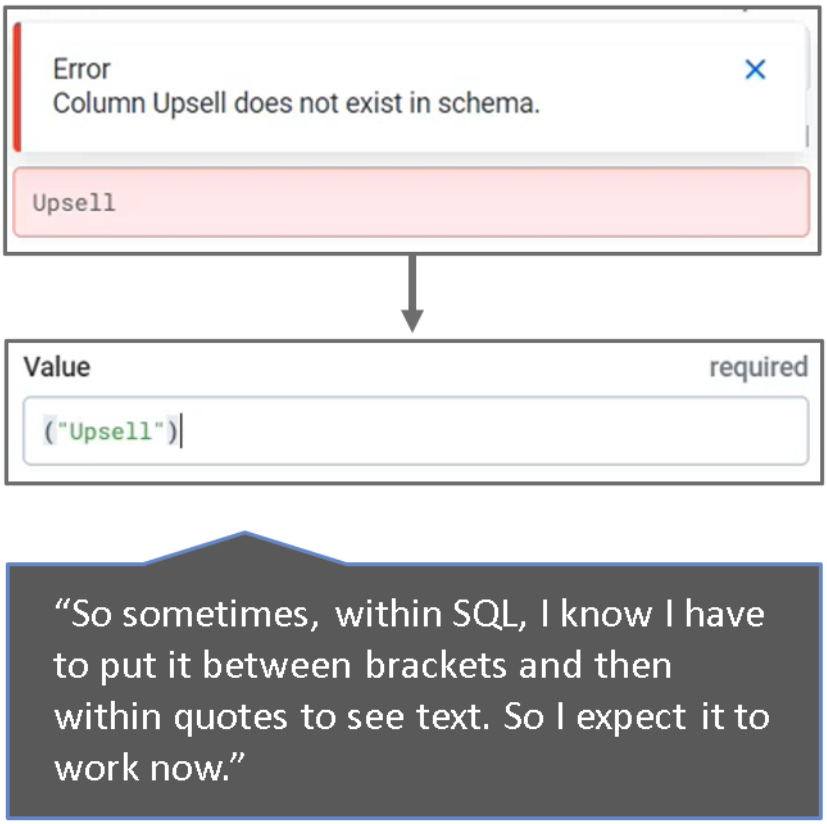


Fig. 4.5.: The participant with SQL knowledge managed to recover from the highly severe planning error of using the wrong syntax for the input field.

In addition, there is another quote illustrating that SQL knowledge helped recognize a hint for solving the same problem. This was because the participant noticed a column label that indicated a string data type. In their words:
'I also saw a column that was ABC. So it was a string type. And I know from SQL that if there is a text. It's an ABC type that it should be within quotes.'
It could be assumed that other participants' lack of SQL knowledge may, therefore, be the reason why they did not perceive the label 'ABC' as useful knowledge.
Let us now move on to how the four participants that use PowerBI daily or weekly were influenced by their experience compared to non-PowerBI users. The latter experienced 4,5 planning issues on average, while PowerBI users only experienced two on average. Therefore, it can be said that experience with PowerBI was connected to the occurrence of planning issues. It is important to mention that the participant with SQL experience did not report using PowerBI. Furthermore, PowerBI was not mentioned in the transcripts so no additional qualitative insights could be found.

5 Discussion, Limitations and Further Work

Prior studies in BI have noted the need for casual users to achieve higher levels of self-service when leveraging data for reporting. In reviewing the literature, no data was found on the usability of data preparation tools. Furthermore, little was found on the association between casual users' technical skill and their ability to conduct data structuring tasks.

5.1 RQ 1: Discussing the usability problems

The first question in this study sought to determine what usability problems casual users face during data structuring tasks in data preparation tools. Looking at the performance of data structuring tasks, participants had the greatest difficulty with aggregation tasks. The frequent failure of the aggregation task indicates that not all data structuring tasks have the same difficulty for casual users. Therefore, there may be data structuring tasks that casual users can already conduct without the help of IT, which can increase their level of self-service.

An interesting finding was that apart from medium-severe problems in the translation category, the most severe problem was in the planning category. It arose from the system's conceptual model not matching the user's mental model. This so-called bridging of the gap between both sides is a key theme of Donald Norman's theory of action as described in the related work chapter [Nor87]. Based on the presented results, it seems as if usability problems in the planning category are more severe than problems in the translation category, and assessment problems are the least severe. Neither Donald Norman nor the researchers that published research about the UAF mention a hierarchy of usability problems within the interaction cycle's phases [And+01; And+00; Kha+11; Nor87].

The results from the SUS highlighted participants' struggle with the ease of use and the learnability of Trifacta Wrangler. Participants with a low SUS score rated Trifacta Wrangler as having poor learnability. Based on this, learnability could be seen as an important aspect of a data preparation tool's usability. This aligns with the BI usability criteria reviewed in the related work. Jooste et al. and Smuts et al. both included learnability as a criterium which is not the case for the 10 usability heuristics by Nielsen [JVM13; SSC15; Nieb]. Eckerson's research also emphasizes the importance of learnability with his characterization of casual users: "despite hours of training, casual users quickly forget how to use BI tools" [Eck12].

5.2 RQ 2: Discussing the influence of technical skill on usability problems

The second question in this research was how technical skill - equated with the use of and experience with data operation technologies and tools - affects the performance of data structuring tasks. An unexpected finding was that participants failed tasks that they would reportedly be able to solve in Excel. In contrast, participants in the 2011 Wrangler study were faster at conducting data operations in the data preparation tool than in Excel [Kan+11]. In the referenced study, however, they received training on performing transformations before conducting the tasks. Participants in this study did not receive training and were allowed to discover the environment themselves after watching a video introducing Trifacta Wrangler. Thus, software training could have influenced participants' interactions with Trifacta Wrangler positively. However, in practice training adds to the cost of BI tools [Eck08] which is why it would be in the interest of casual users not to require training. Applying human-centered design during the software development cycle as done in this evaluation study can help reduce training costs [ISO19]. Therefore, it can be argued that this study's findings are highly relevant to the field of BI. Furthermore, this discussion adds to the importance of learnability as a usability criterium as discussed in the previous section.

The current study found that SQL knowledge improves task performance and reduces the number of severe usability problems occurring. It can be said that technical knowledge and, specifically, experience with concepts of programming languages influenced the usability problems in this study. This is in line with Hameed et al. who postulated that "A typical domain-expert cannot be expected to formulate often intricate regular expressions" [HN20].

In this case, the input code could be considered simpler than a regular expression. Convertino et al. similarly emphasized the dependence of casual users without programming knowledge on power users [CE17]. The lack of technical skill also relates to the topic of data democratization introduced in the section scientific and social relevance of the introduction. Two of data democratization's enablers are the development of data and analytic skills, and self-service analytics tools [LLF21]. Since casual users can not be expected to develop their technical skill it is the responsibility of self-service tools to support them with their data operations. Therefore, data preparation tools must not only be usable for power users but also for casual users. This is in line with the conclusion drawn by Convertino et al., that there is a need for platforms that integrate multiple users with different technical skills in an organization [CE17].

Moving on to PowerBI, its experience led participants to encounter fewer issues in the planning category. Thus, beyond the scope of the tools identified in the related works, there seem to be more tools that can influence the technical skill of casual users. This is in line with Jakob's Law on website user experience, which states that users spend a lot of time in other tools, influencing their interactions in any tool that is being investigated [Nie00]. Thus, identifying interaction patterns from other tools casual users use could help improve the usability of data preparation tools. This knowledge could be applied to simplify the programming of data queries or incorporate more user-centred wording and thus addressing one of the identified medium-severe usability issues.

5.3 Limitations and implications for further research

The major limitation of this study is that the transcript was coded by one researcher, and the usability problems were partly categorized by an additional researcher. This was due to resource scarcity and the difficulty finding HCI experts experienced with the User Action Framework. However, according to Andre et al., the UAF maintains strong reliability with one coder for the first two levels of the problem classification tree [And+00]. Therefore, the results reporting on the categories of Planning, Translation, and Assessment, including their first subcategory can still be considered meaningful to the field of BI usability research.
The generalizability of the results is also influenced by the following limitations: One participant's screen recording was lost due to technical issues resulting in only using verbalizations from the audio recording to identify their usability problems. Furthermore, the study's sample size influences the System Usability Scale's reliability. According to Tullis et al., the SUS requires at least 12 participants [TS04]. In their study, they postulate that a sample size of 8 participants, as used in this study, leads to less than 50% of conclusions reached from the SUS being correct.
However, the conclusions of this study were mainly drawn based on the UAF framework, and the usability problems seemed to converge towards the end of the experiments. Despite this, it would be useful to repeat the study with a larger sample size and more experts for encoding the data. Furthermore, conducting a study outside of a controlled environment may inform what the influence of certain usability problems on a real project is as intended by Khajouei's User Action Framework [Kha+11].

Concerning the study design, the focus laid on a subset of the data structuring tasks mentioned by Rattenburry [Rat+]. Future work should investigate task performance for *pattern extraction*, *complex structure extraction*, *combining multiple record fields*, and *pivots*. This could lead to identifying other data operations that casual users can conduct independently and reveal additional usability problem categories than the ones identified in this study. Based on another study design decision, the task descriptions were provided in natural language. It would be interesting to see if the usability problems or participants' task performance were different if they were given a visualization of the desired outcome data instead of a task description.
Lastly, to the best of the author's knowledge, this was one of the first studies on the usability of English BI tools conducted in a Dutch business context. As can be seen in figure 4.1 the code *langaugeBarrier* only consists of two references indicating this limitation to have had a small impact.

However, future studies have the opportunity to make adjustments and focus on the interaction of participants that speak English as their second language with English data preparation tool interfaces.

6 Conclusion

This research aimed to gain insights into how casual users can reach more independence from power users when using data for decision-making. The paper presents findings on the type of usability problems casual users face when conducting data structuring tasks in Trifacta Wrangler. In addition, it shows how technical skill influences the occurrence of specific problems. Eight participants from teamITG participated in the study and completed a survey, including the system usability scale next to participating in the usability test.

The results show a tendency toward specific usability problem categories from the User Action Framework. Most of the 68 usability problems were distributed between the planning and translation categories, and two problems were in the assessment category. The only problem with high severity, and thus leading to task failure most often, was in the planning category and relates to a mismatch between the mental model of the user and the conceptual model of the system. This was most apparent during the aggregation task, which only two participants completed. In other tasks, participants were often unfamiliar with the programming syntax which is in line with related works' findings that casual users struggle with technologies such as regular expressions. In this study, only the participant experienced with SQL recovered from the severe problem due to their knowledge of programming concepts. Excel skills influenced participants feeling of familiarity positively; however, contrary to prior research, the Trifacta Wrangler did not lead participants to easily conduct data operations. Participants not receiving training before the test could have impacted this finding.

Concluding, for casual users within business intelligence (BI) to conduct data structuring tasks independently, data preparation tools must improve the match between the system's conceptual model and the casual user's mental model. Thereby, casual users' experience with and use of other data operating tools, especially Excel, must be considered. Future research should enrich this study's findings about by conducting quantitative studies on casual users' technical skill within the Dutch BI context. Qualitative studies with casual users should investigate which types of usability problems occur during other data structuring tasks and use multiple evaluators to code the results. Finally, future research on BI tools could investigate how to improve their learnability - as a usability criterium - for casual users.

Bibliography

[AS] Paul Alpar and Michael Schulz. “Self-Service Business Intelligence”. In: () (cit. on pp. 1, 2, 4, 14).

[AT] Marcel Altendeitering and Martin Tomczyk. “A Functional Taxonomy of Data Quality Tools: Insights from Science and Practice”. In: () (cit. on p. 2).

[And+00] Terence S. Andre, Steven M. Belz, Faith A. McCreary, and H. Rex Hartson. “Testing a framework for reliable classification of usability problems”. In: *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, ’Ergonomics for the New Millennium’*. 2000 (cit. on pp. 24, 49, 52).

[And+01] Terence S. Andre, H. Rex Hartson, Steven M. Belz, and Faith A. McCreary. “User action framework: a reliable foundation for usability engineering support tools”. In: *International Journal of Human Computer Studies* 54.1 (2001), pp. 107–136 (cit. on p. 49).

[Bar20] Carol M. Barnum. *Usability Testing Essentials: Ready, Set. . . Test!* 2020 (cit. on pp. 33, 65).

[Bro96] John Brooke. “SUS: A ’Quick and Dirty’ Usability Scale”. In: *Usability Evaluation In Industry*. 1996, pp. 4–7 (cit. on pp. 31, 66).

[Bru] Anders Bruun. *Training tool for the UAF* (cit. on p. 35).

[CCS12] Hsinchun Chen, Roger H.L. Chiang, and Veda C. Storey. “Business intelligence and analytics: From big data to big impact”. In: *MIS Quarterly: Management Information Systems* 36.4 (2012) (cit. on p. 1).

[CE17] Gregorio Convertino and Andy Echenique. “Self-service data preparation and analysis by business users: New needs, skills, and tools”. In: *Conference on Human Factors in Computing Systems - Proceedings* Part F127655 (May 2017), pp. 1075–1083 (cit. on pp. 2, 3, 5, 15, 16, 20, 31, 32, 51).

[DA17] Mohammad Daradkeh and Radwan Moh d. Al-Dwairi. “Self-service business intelligence adoption in business enterprises: The effects of information quality, system quality, and analysis quality”. In: *International Journal of Enterprise Information Systems* 13.3 (July 2017), pp. 65–85 (cit. on p. 1).

[] *Data Wrangling Software and Tools - Trifacta* (cit. on pp. 21, 22).

[Dum02] Joseph S Dumas. “User-based evaluations”. In: *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. 2002, pp. 1093–1117 (cit. on p. 27).

[Eck11] Wayne Eckerson. “Big Data Analytics: Profiling the Use of Analytical Platforms in User Organizations”. In: (2011) (cit. on pp. 2, 7–9, 12, 13).

[Eck12] Wayne Eckerson. “Business-Driven BI using new Technologies to Foster self-service Access to insights”. In: (2012) (cit. on pp. 20, 50).

[Eck09] Wayne Eckerson. “Self-Service BI”. In: (2009) (cit. on p. 2).

[Eck08] Wayne w. Eckerson. *Pervasive Business Intelligence - Techniques and Technologies to Deploy BI on an Enterprise Scale*. Tech. rep. TDWI, 2008 (cit. on pp. 2, 13, 14, 19, 50).

[Fan+19] Mingming Fan, Jinglan Lin, Christina Chung, and Khai N. Truong. “Concurrent think-aloud verbalizations and usability problems”. In: *ACM Transactions on Computer-Human Interaction* 26.5 (2019) (cit. on p. 38).

[GH15] Amir Gandomi and Murtaza Haider. “Beyond the hype: Big data concepts, methods, and analytics”. In: *International Journal of Information Management* 35.2 (2015), pp. 137–144 (cit. on p. 10).

[HN20] Mazhar Hameed and Felix Naumann. “Data Preparation: A Survey of Commercial Tools”. In: (2020) (cit. on pp. 3, 50).

[HAT18] Ayad Hameed Mousa, Heba Adnan Raheem, and Nibras Talib Mohammed. “An Evaluation Instrument (Q-U) for Measuring the usability of Business Intelligence Application”. In: *International Journal of Engineering & Technology* 7.4.19 (2018) (cit. on p. 25).

[HHK18] Joseph M Hellerstein, Jeffrey Heer, and Sean Kandel. “Self-Service Data Preparation: Research to Practice”. In: *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* (2018) (cit. on pp. 2, 3, 11).

[IKS19] Igor Ilin, Anastasii Klimin, and Anton Shaban. “Features of Big Data approach and new opportunities of BI-systems in marketing activities”. In: *E3S Web of Conferences*. Vol. 110. 2019, p. 2054 (cit. on p. 1).

[IW11] Claudia Imhoff and Colin White. “Self-service business intelligence: Empowering Users to Generate Insights”. In: *TDWI best practices report* (2011) (cit. on pp. 1, 3).

[ISO19] ISO. *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*. 2019 (cit. on p. 50).

[Jää10] Riitta Jääskeläinen. “Think-aloud protocol”. In: 2010 (cit. on p. 30).

[JAC15] Björn Johansson, Dogan Alkan, and Robin Carlsson. “Self-Service BI does it Change the Rule of the Game for BI Systems Designers”. In: (2015) (cit. on p. 13).

[JVM13] Chrisna Jooste, Judy Van Biljon, and Jan Mentz. “Usability evaluation guidelines for Business Intelligence applications”. In: *ACM International Conference Proceeding Series*. 2013 (cit. on pp. 25, 50).

[Jor+96] PW Jordan, B Thomas, IL McClelland, and B Weerdmeester. *Usability Evaluation In Industry*. 1996 (cit. on p. 29).

[Kan13] Sean Kandel. "Interactive Systems for Data Transformation and Assessment". PhD thesis. Stanford University, 2013 (cit. on p. 21).

[Kan+11] Sean Kandel, Andreas Paepcke, Joseph Hellerstein, and Jeffrey Heer. "Wrangler: Interactive Visual Specification of Data Transformation Scripts". In: (2011) (cit. on pp. 23, 50).

[Kan+12] Sean Kandel, Andreas Paepcke, Joseph M Hellerstein, and Jeffrey Heer. "Enterprise Data Analysis and Visualization: An Interview Study". In: (2012) (cit. on pp. 2, 15, 16, 18–20, 31, 32).

[Kee+99] Susan L. Keenan, H. Rex Hartson, Dennis G. Kafura, and Robert S. Schulman. "The Usability Problem Taxonomy: A Framework for Classification and Analysis". In: *Empirical Software Engineering 1999 4:1* 4.1 (Mar. 1999), pp. 71–104 (cit. on p. 24).

[Kha+11] R. Khajouei, L. W.P. Peute, A. Hasman, and M. W.M. Jaspers. "Classification and prioritization of usability problems using an augmented classification scheme". In: *Journal of Biomedical Informatics* 44.6 (Dec. 2011), pp. 948–957 (cit. on pp. 24, 37, 49, 52, 67).

[LLF21] Hippolyte Lefebvre, Christine Legner, and Martin Fadler. "Data democratization: toward a deeper understanding Open Data to Business-Discovery, Integration and Use of Open Data in Business Environments View project Don't Guess, Simulate! Understanding User Preferences for Cloud Services View project". In: (2021) (cit. on pp. 4, 51).

[LVS] Christian Lennerholt, Joeri Van Laere, and Eva Söderström. "Implementation Challenges of Self Service Business Intelligence: A Literature Review". In: () (cit. on p. 19).

[Lew14] James R. Lewis. *Usability: Lessons Learned. and Yet to Be Learned*. 2014 (cit. on pp. 28, 29).

[Lit17] Cinny Little. "The Forrester Wave™: Data Preparation Tools". In: (2017) (cit. on p. 20).

[Log14] Logi Analytics. *2014 State of self-service BI report*. Tech. rep. Logi Analytics, 2014 (cit. on p. 16).

[MW 94] J.A.C Sandberg M.W. van Someren Y.F. Barnard. *The think aloud method: a practical approach to modelling cognitive process*. Vol. 31. 6. 1994 (cit. on pp. 29, 30).

[MN93] Robert Mack and Jakob Nielsen. "Usability inspection methods". In: *ACM SIGCHI Bulletin* 25.1 (1993) (cit. on p. 37).

[Mic+20] Sven Michalczyk, Mario Nadj, Alexander Maedche, and Christoph Gröger Bosch. “A State-Of-The-Art Overview and Future Research Avenues of Self-Service Business Intelligence and Analytics Design Science Research Methodology View project Designing Interactive Crowd-Feedback Systems View project”. In: (2020) (cit. on pp. 2, 3, 19, 20).

[Mic] Microsoft. *What is Power BI | Microsoft Power BI* (cit. on p. 39).

[MD08] Rolf Molich and Joseph S. Dumas. “Comparative usability evaluation (CUE-4)”. In: *Behaviour and Information Technology* 27.3 (2008) (cit. on p. 29).

[Mol+04] Rolf Molich, Meghan R. Ede, Klaus Kaasgaard, and Barbara Karyukin. “Comparative usability evaluation”. In: *Behaviour and Information Technology* 23.1 (2004) (cit. on p. 29).

[Nie00] Jakob Nielsen. *End of Web Design*. 2000 (cit. on p. 51).

[Niea] Jakob Nielsen. “FINDING USABILITY PROBLEMS THROUGH HEURISTIC EVALUATION”. In: () (cit. on p. 27).

[Nieb] Jakob Nielsen. “Heuristic Evaluation Ten Usability Heuristics”. In: () (cit. on pp. 24, 27, 50).

[Nie94] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann, 1994 (cit. on pp. 24, 30, 37).

[NL93] Jakob Nielsen and Thomas K. Landauer. “Mathematical model of the finding of usability problems”. In: *Conference on Human Factors in Computing Systems - Proceedings*. 1993 (cit. on pp. 27, 29).

[NMB90] Jukob Nielsen, Rolf Molich, and Jn@neuvml Bitnet Denmark. “CHI 90 Procee&qs HEURISTIC EVALUATION OF USER INTERFACES”. In: (1990) (cit. on p. 27).

[Nor87] Donald A. Norman. “Cognitive Engineering”. In: *Interfacing thought: cognitive aspects of human-computer interaction*. 1987, pp. 325–336 (cit. on pp. 30, 49).

[Ott] OtterAI. *Otter.ai - Voice Meeting Notes & Real-time Transcription* (cit. on p. 35).

[PLB] Jens Passlick, Benedikt Lebek, and Michael H Breitner. “A Self-Service Supporting Business Intelligence and Big Data Analytics Architecture”. In: () (cit. on pp. 7–9, 12, 14).

[PW] Gloria Phillips-Wren and Hugh J Wat. “Business Analytics in the Context of Big Data: A Roadmap for Research”. In: () (cit. on pp. 2, 14–16).

[Pre+11] Jenny Preece, Helen Sharp, Rogers Yvonne, Yvonne Rogers, and Jenny Preece. “Interaction Design: Beyond Human-Computer Interaction, 3rd Edition”. In: *Book* 11 (2011) (cit. on p. 27).

[Qua] Qualtrics. *Marktforschung, Umfrage & Erlebnismanagement Software I Qualtrics* (cit. on p. 31).

[Rat+] Tye Rattenbury, Joe Hellerstein, Jee Rey Heer, Sean Kandel, and Connor Carreras. “Principles of Data Wrangling”. In: () (cit. on pp. 11, 12, 14, 18, 19, 31, 52).

[Ros22] Maria Rosala. *Checklist for Moderating a Usability Test*. May 2022 (cit. on pp. 33, 65).

[RH21] Christopher M. Rosett and Austin Hagerty. “Introducing HR Analytics with Machine Learning”. In: *Introducing HR Analytics with Machine Learning* (2021) (cit. on p. 20).

[Sav14] Alexandr Savinov. “ConceptMix: Self-service analytical data integration based on the concept-oriented model”. In: *DATA 2014 - Proceedings of 3rd International Conference on Data Management Technologies and Applications*. 2014 (cit. on p. 2).

[Sie+13] David Siegel, Alex Sorin, Michael Thompson, and Susan Dray. “Fine-tuning user research to drive innovation”. In: *Interactions* 20.5 (2013), pp. 42–49 (cit. on pp. 18, 19).

[SSC15] Martin Smuts, Brenda Scholtz, and Andre Calitz. “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 - SAICSIT ’15*. New York, New York, USA: ACM Press, 2015, pp. 1–15 (cit. on pp. 25, 50).

[Sta] Stanford. *Data Wrangler* (cit. on p. 21).

[Ste+] Darko Stefanović, D Narandžić, T Lolić+, D Stefanović+, and S Ristić+. “The Challenge of an Extraction-Transformation-Loading Tool Selection”. In: () (cit. on p. 21).

[Sto16] David Stodder. “Improving Data Preparation for Business Analytics Applying Technologies and Methods for Establishing Trusted Data Assets for More Productive Users BEST PRACTICES REPORT Q3 2016”. In: (2016) (cit. on p. 2).

[Sto15] David Stodder. “Visual Analytics for Making Smarter Decisions Faster Applying Self-Service Business Intelligence Technologies to Data-Driven Objectives”. In: *TDWI Best Practices Report* (2015) (cit. on p. 1).

[14] “The design of everyday things”. In: *Choice Reviews Online* 51.10 (2014) (cit. on p. 30).

[TP09] Christian Thomsen and Torben B. Pedersen. “A survey of open source tools for Business Intelligence”. In: *International Journal of Data Warehousing and Mining* 5.3 (2009) (cit. on p. 33).

[Tri] Alteryx Trifacta. *Getting Started with Trifacta - YouTube* (cit. on p. 33).

[TS04] Thomas S Tullis and Jacqueline N Stetson. “A Comparison of Questionnaires for Assessing Website Usability”. In: *Proceedings of UPA 2004 Conference*. 2004 (cit. on p. 52).

[UIE] UIE. *Eight is Not Enough — UX Articles by UIE* (cit. on p. 27).

[VDS03] Maaike J. Van Den Haak, Menno D.T. De Jong, and Peter Jan Schellens. "Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue". In: *Behaviour and Information Technology* 22.5 (2003) (cit. on pp. 30, 35, 38).

[VES86] M. Venkatesan, K. Anders Ericsson, and Herbert A. Simon. "Protocol Analysis: Verbal Reports as Data". In: *Journal of Marketing Research* 23.3 (1986) (cit. on p. 30).

[Wat09] Hugh Watson. "Business Intelligence: Past, Present and Future. Top Concerns of BI and Analytics Managers View project". In: (2009) (cit. on pp. 13, 20).

[Wat15] Hugh J Watson. "Data Lakes, Data Labs, and Sandboxes". In: *Business Intelligence Journal* 20.1 (2015) (cit. on pp. 15, 16).

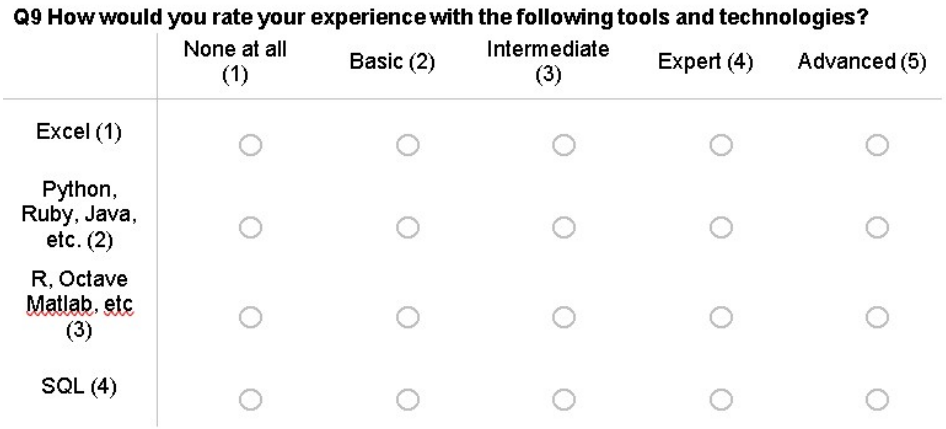
[ZSV17] Ehtisham Zaidi, Rita L Sallam, and Shubhangi Vashisth. "Market Guide for Data Preparation". In: *Gartner* (2017) (cit. on pp. 20, 21).

[ZSV] Ehtisham Zaidi, Rita L Sallam, and Shubhangi Vashisth. "Market Guide for Data Preparation". In: () (cit. on p. 2).

[ZG18] Jing Zeng and Keith W Glaister. "Strategic Organization". In: 16.2 (2018), pp. 105–140 (cit. on p. 4).

Appendix

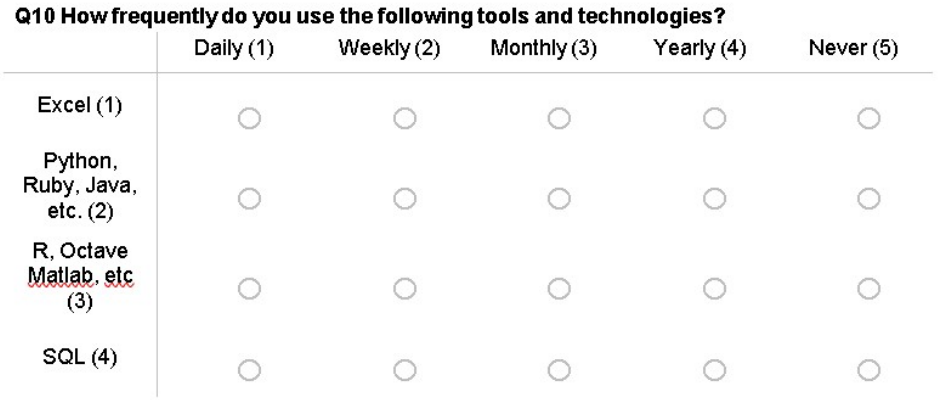
A



Q9 How would you rate your experience with the following tools and technologies?

	None at all (1)	Basic (2)	Intermediate (3)	Expert (4)	Advanced (5)
Excel (1)	○	○	○	○	○
Python, Ruby, Java, etc. (2)	○	○	○	○	○
R, Octave Matlab, etc (3)	○	○	○	○	○
SQL (4)	○	○	○	○	○

Fig. A.1.: The survey question inquiring about participants' experience with tools and technologies



Q10 How frequently do you use the following tools and technologies?

	Daily (1)	Weekly (2)	Monthly (3)	Yearly (4)	Never (5)
Excel (1)	○	○	○	○	○
Python, Ruby, Java, etc. (2)	○	○	○	○	○
R, Octave Matlab, etc (3)	○	○	○	○	○
SQL (4)	○	○	○	○	○

Fig. A.2.: The survey question inquiring about participants' use of tools and technologies

Q11 How familiar are you with the following data operations?

	Not familiar at all (12)	Moderately familiar (13)	Very familiar (14)	Click to write Scale Point 4 (17)
Text Extraction (1)	○	○	○	○
Filtering Records and Fields (2)	○	○	○	○
Aggregations (3)	○	○	○	○

Fig. A.3.: The survey question inquiring about participants' familiarity with data structuring operations

Q12 Which other tools or technologies do you use when working with data and how frequently do you use them?

	Daily (1)	Weekly (2)	Monthly (3)	Yearly (4)	Never (5)
Input Tool 1 (1)	○	○	○	○	○
Input Tool 2 (2)	○	○	○	○	○
Input Tool 3 (3)	○	○	○	○	○

Fig. A.4.: The survey question inquiring about participants' use of tools and technologies not mentioned in Q9 and Q10

Moderator's checklist

Before participant arrives

○□ Make sure product is loaded properly and ready for first scenario
○□ Make sure survey is loaded properly and ready to be completed
○□ Make sure phone and microphone are properly positioned

Welcome

○□ Introduce yourself, thank participant for having an interest in participating
○□ Offer refreshment
○□ Escort participant to evaluation room
○□ Ask participant to sit at the desk
○□ Sit beside the participant

Consent form, pre-test questionnaire, instructions

○□ Show participant the location of cameras, phone, microphone
○□ Explain the purpose of the test
○□ Go over consent form, allow time to read and sign; if this has been done already, ask participant if he/she is comfortable with being recorded
○□ Explain that there are observers who are very interested in learning from the participant about his/her experience
○□ Ask for questions, concerns
○□ Give pre-test questionnaire

Instructions

○□ Explain process of using scenarios, one at a time, while participant thinks out loud
○□ Review how think-out-loud process works, with examples
○□ Demonstrate how to use the phone to call the help desk or to indicate completion of a scenario
○□ Explain that after each scenario, there will be a quick questionnaire to complete, then the next scenario

After each scenario

○□ Offer plenty of reassurance, especially when tasks prove difficult
○□ Give feedback on the quality of the think-out-loud procedure; if necessary, encourage more feedback from participant by reviewing the process again, with examples
○□ Ask participant to clarify any thoughts or actions
○□ Set up product at starting point for next scenario, if needed

After completion, post-test questionnaire

○□ Give post-test questionnaire (or whatever feedback mechanisms are being used)
○□ Thank for experience
○□ Provide stipend for participation (or direct participant to office for payment)

Fig. A.5.: Moderator's checklist used as a preparation for the usability tests based on two publications by Barnum and Rosala [Bar20; Ros22]

Please score the system according to the following 10 items with one of five responses that range from Strongly Agree to Strongly disagree:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I think that I would like to use this system frequently. (1)	○	○	○	○	○
I found the system unnecessarily complex. (2)	○	○	○	○	○
I thought the system was easy to use. (3)	○	○	○	○	○
I think that I would need the support of a technical person to be able to use this system. (4)	○	○	○	○	○
I found the various functions in this system were well integrated. (5)	○	○	○	○	○
I thought there was too much inconsistency in this system. (6)	○	○	○	○	○
I would imagine that most people would learn to use this system very quickly. (7)	○	○	○	○	○
I found the system very cumbersome to use. (8)	○	○	○	○	○
I felt very confident using the system. (9)	○	○	○	○	○
I needed to learn a lot of things before I could get going with this system. (10)	○	○	○	○	○

Fig. A.6.: The part of the survey containing the System Usability Scale [Bro96]

Q15 Task A:

Given the dataset SALES OPPORTUNITIES.
Perform an extraction on the column "CLOSEDATE" to extract the Day into a new column.

Q16 Task B:

Given the dataset SALES OPPORTUNITIES.
Filter the dataset to only include records of the type Upsell in the column "TYPE".

Q17 Task C:

Given the dataset SALES OPPORTUNITIES.
Looking at each lead source category in the column "LEADSOURCE". For each leadsource aggregate the sum of the number of leads contained in the column "AMOUNT" into a new column

Fig. A.7.: The three data structuring tasks that participants were instructed to perform during the usability test

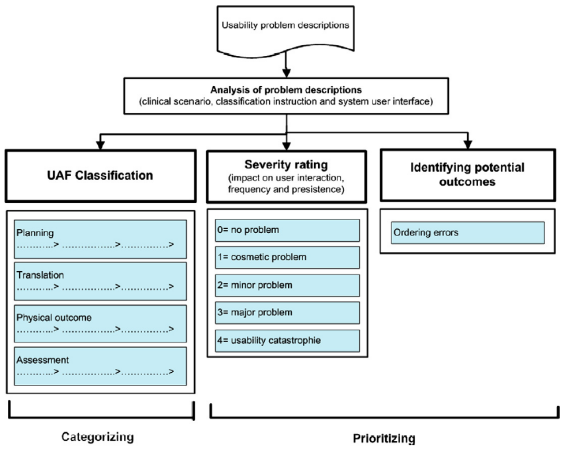
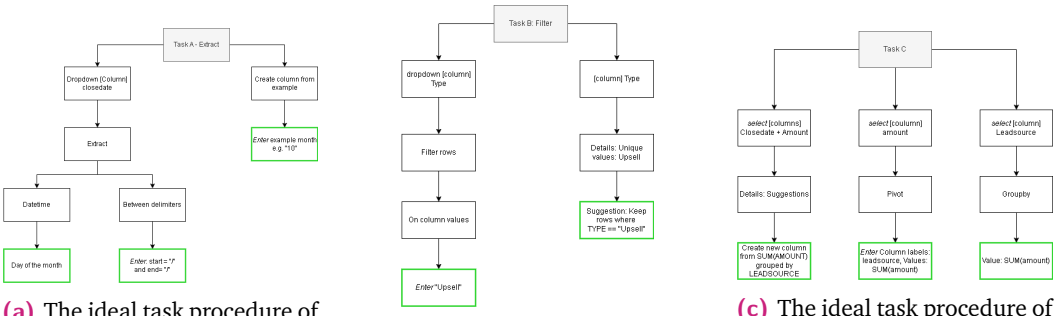


Fig. A.8.: The User Action Framework (UAF) as a method of classifying usability problem descriptions according to Khajouei et al. [Kha+11]



(a) The ideal task procedure of the extraction task

(b) The ideal task procedure of the filtering task

(c) The ideal task procedure of the aggregation task

Fig. A.9.: The ideal task procedures of the data structuring tasks used to identify usability problems

Fig. A.10.: Included for better readability: The first part of the User Action Framework phases including the subcategories and their specific codes

Fig. A.11.: Included for better readability: The second part of the User Action Framework phases including the subcategories and their specific codes

	Participant ID	1	2	3	4	5	6	7	8	
Task	A.Extraction	T3b P3a A tb2 P3a T1b1 -21, 29, 1A	T3b	T1h	T1h P1a T3a P3a T1b1	T2b	T3f T3b P3a T1b1 P3a	T3f P3a(reason for task failure) T1b1	T1a A tb2 T3b	
	B.Filtering	P3a	T2ar1 T3 T3f (2nd) T3a		none	P3a T1b1 T3j	P1a P3a T2ar1 P1h1 T1b1	P3a(reason for task failure) T1b1 T3h	P3a	
	C.Aggregation	T1a P3a(reason for task failure)	T1b1 P1a(reason for task failure) P4 A1c2 T1b1 (2nd occurrence)	T1a T1a (2nd)	T1b1 P1a(reason for task failure) T1a T1a	T1a(reason for task failure)	P4 P3a P1a(reason for task failure)	P3a(4 times) (reason for task failure)	T3a(2times) T3f	
SUSscore		32.5	67.5	92.5	47.5	32.5	35	50	57.5	
Department		Sales	Marketing	Sales	Client Services	Sales	Client Services	Marketing	Client Services	
Gender		Male	Male	Male	Female	Male	Male	Male	Female	
Age		45-54	18-24	25-35	35-44	35-44	35-44	35-44	25-35	
Familiarity with data operations (1-5)	text extraction	1	1	1	2	1	2	1	1	1
	Filtering	2	1	2	3	3	2	2	2	2
	Aggregations	2	1	2	1	1	1	1	2	2
Experience (1-5)	Excel	3	3	3	3	4	3	1	1	4
	Python, RUby, Java	1	1	1	1	1	1	1	1	1
	R,Octave, Matlab	1	1	1	1	1	1	1	1	2
Frequency of use (1-5)	SQL	5	4	4	5	5	5	4	4	3
	Excel	4	4	4	5	5	5	4	4	3
	Python, RUby, Java	1	1	1	1	1	1	1	1	1
	R,Octave, Matlab	1	1	1	1	1	1	1	1	1
	SQL	1	1	1	1	1	1	1	1	1
	Power BI	4	N/A	5	5	4	N/A	N/A	N/A	N/A

Task performance: Task completed / Task failed

Fig. A.12.: The survey results with the usability problem codes assigned to each task

Problem ID	Task	Participant ID	Description	Classification
1	A	2	The difference between the extraction options "day of the month" and "day of the week" are not clear	Translation -> Content and Meaning -> Completeness and sufficiency of meaning
2	B	2	The suggestions were not noticed even though they were accurate	Translation -> Presentation -> Perceptual Issues -> Noticability
3	C	2	It is unknown that there is a functionality to select two columns at the same time and receive suggestions for them	Translation -> Existence -> Existence of a cognitive affordance -> to show the way
4	C	2	Participants don't know how to conceptualize the aggregation and fail to plan their activities. (They do not recognize relevant terminology for it in the system)	Planning -> Goal decomposition ->User unable to establish sequence of tasks to accomplish goal
6	C	2	Writing of code in a pragmatic manner that involves table values as references rather than column variables. This way users type in their code close to what the task description in their head is 'count number of websites' I wanna type in like website and then it gives me the amount	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
7	C	2	Interpreting the output of a groupby and count() and not understanding what it means	Planning -> Users's inability to determine what to do next
8	C	2	Feedback states invalid formula, but does not further detail the reason nor does it suggest a recovery	Assessment -> Issues of feedback -> Content and Meaning -> Completeness and sufficiency of meaning
9	C	2	Managed to group them but failed to find a way to add them up according to their count in the other collumn	Translation -> Existence -> Existence of a cognitive affordance to show the way.
10	A	3	Trying to rename column by clicking on title header in the sidebar and starting to type	Translation>Existence>Existence of a way-> Missing feature
11	B	3	Filter rows confusion about the terminology as it refers to rows not columns	Translation -> Content and Meaning -> User centeredness of wording, design of cognitive affordance content
12	B	3	In excel there is an option to filter that you click on but here it is a dropdown with options that the user did not concider. "Wasnt the type of filter that I needed" this lead to cancellation and looking for another way	Translation -> Content and Meaning -> User centeredness of wording, design of cognitive affordance content
13	B	3	Double click was not registered by the system to select the row which made participant unaware of the possibility to select it	Translation -> Preferences and efficiency -> Alternative way(s) to do task, step
14	C	3	Participant is looking for a button to create a new column	Translation -> Existence -> Existence of a way-> Missing feature
15	C	3	Participant wants to enrich a pivot table that was created with its original dataset (Excel)	Translation -> Existence -> Existence of a way-> Missing feature
16	A	4	Click on a single cell to receive a suggestion for an action on that column based on the function of receiving suggestions from clicking on the statistics bars	Translation -> Existence -> Existence of a way-> Missing feature
17	A	4	User looked for suggestions as help to resolve the task	Planning -> Goal decomposition ->User unable to establish sequence of tasks to accomplish goal
18	A	4	Clicked on statistical bars to receive general suggestions for all data points but received specific ones to a single data point	Translation -> Content and meaning -> Clarity, precision, predictability of meaning
19	A	4	User tries to enter a single slash without quotations as a delimiter value for splitting	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
20	A	4	Feedback not aiding error recovery after entering forward-slash as a delimiter value. The feedback is a suggestion that does not fit the user's intention and does not provide smart ideas	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.

Fig. A.13.: The first part of the usability problem inventory which is based on analyzing the observation and verbalization recordings

21	C	4	Did not notice the way to hide columns	Translation -> Existence -> Existence of a cognitive affordance -> Existence of a cognitive affordance to show the way.
22	C	4	User looked for suggestions as help to resolve the task	Planning -> Goal decomposition ->User unable to establish sequence of tasks to accomplish goal
23	C	4	Enter formula (based on excel knowledge) into create column by example. This is not supported by the system	Translation -> Existence -> Existence of a way-> Missing feature
24	C	4	Participants don't know how to conceptualize the aggregation and fail to plan their activities. They do not recognize relevant terminology for it in the system	Planning -> Goal decomposition ->User unable to establish sequence of tasks to accomplish goal
25	C	4	Drag and drop elements like they would in excel with pivot tables	Translation -> Existence -> Existence of a way-> Missing feature
26	A	5	A tooltip about formula input blocked the users view and access of a button they wanted to interact with	Translation -> Presentation -> Layout and grouping
27	B	5	Filter rows by matching exact value: upsell. Did not add the quotations	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
28	B	5	Feedback not aiding error recovery after entering string without quotations. The feedback is does not provide smart ideas as to the correct action.	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.
29	B	5	Terminology 'Filter' does not fit with the 'delete row' option of the same process. From the user's perspective they are both different manipulations of data.	Translation -> Content and Meaning -> Consistency and compliance of cognitive affordance meaning
30	C	5	Participant wants to create a new cell, row or column to sum up all values of a column as they would in an excel sheet	Translation -> Existence -> Existence of a way-> Missing feature
31	A	6	Search logo was interpreted to enable searching for columns. Then the header implied search for transformations.	Translation -> Content and meaning -> User centeredness of wording, design of cognitive affordance content
32	A	6	The difference between the extraction options "day of the month" and "day of the week" are not clear	Translation -> Content and Meaning -> Completeness and sufficiency of meaning
33	A	6	Entering a formula to extract the day, the participant struggled to plan a formula and needed documentation/suggestions to make a choice	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
34	A	6	Error message for formula wrong schema was not helpful for participant to recover or learn	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.
35	A	6	User tries to enter a single slash without quotations as a delimiter value for splitting	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
36	B	6	Participant was going for groupby instead of a filter. *Actually also a way I think!*	Planning -> Goal decomposition -> User unable to establish sequence of tasks to accomplish goal
37	B	6	User tries to enter a single slash without quotations as a delimiter value for splitting	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
38	B	6	Clicks on upsell row smart selector, doesnt notice the suggestions change at first. First suggestion is the one participant is seeking	Translation -> Presentation -> Perceptual issues -> Noticability
39	B	6	When selecting upsell smart selector, the user did not know that he wanted to keep the selected rows *This could also be a user centered meaning issue...*	Planning -> Goal decomposition -> Users ability to determine what to do next
40	B	6	Feedback not aiding error recovery after entering string without quotations. The feedback is does not provide smart ideas as to the correct action.	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.
41	C	6	Participant does not know if count() is what he/she is looking for	Planning -> user and work context
42	C	6	Participant does not know how to type in formula correctly	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
43	C	6	Participant cannot conceptualize an aggregation	Planning -> Goal decomposition ->User unable to establish sequence of tasks to accomplish goal
44	A	7	Extract between positions mistaken for delimiter option	Translation -> Content and meaning -> User centeredness of wording, design of cognitive affordance content
45	A	7	User tries to enter a single slash without quotations as a delimiter value for splitting	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
46	A	7	Error message for formula wrong schema was not helpful for participant to recover or learn	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.
47	B	7	User tries to enter a single slash without quotations as a delimiter value for filtering	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
48	B	7	Error message for formula wrong schema was not helpful for participant to recover or learn	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.

Fig. A.14.: The second part of the usability problem inventory which is based on analyzing the observation and verbalization recordings

49	B	7	Question mark tooltip above value input does not offer help to prevent error in data entry	Translation -> Content and meaning -> Error avoidance
50	C	7	Participant wants to hold down shift and select multiple rows as they would in Excel	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
51	C	7	User tries to enter a string value without quotations to filter values	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
52	C	7	Entering a formula to groupBy but failing due to the syntax	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
53	C	7	Writing of code in a pragmatic manner that involves table values as references rather than column variables. This way users type in their code close to what the task description in their head is 'count number of websites' I wanna type in like website and then it gives me the amount	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
54	A	8	Wants to add a new empty row as they would in excel	Translation -> Existence -> Existence of a way
55	A	8	When an operation is selected on a column that is on furthers right side of the screen that is still visible, then the preview is not visible due to the sidebar popup blocking the original column and the preview opening outside of the visible area.	Assessment -> Issues about feedback -> Presentation -> Layout and Grouping
56	A	8	The difference between the extraction options "day of the month" and "day of the week" are not clear	Translation -> Content and Meaning -> Completeness and sufficiency of meaning
57	B	8	User tries to enter a string value without quotations to filter values	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
58	C	8	user tries to apply hide-preference to multiple selected columns at ones	Translation -> Preferences and efficiency -> Alternative way(s) to do task, step
59	C	8	Tried to click on column and drag it to a desired position in order to move it	Translation -> Preferences and efficiency -> Alternative way(s) to do task, step
60	C	8	User did not consider suggestion as they were cognitively overwhelming	Translation -> Content and meaning -> Relevance of content, meaning OR Mnemonical meaningful cognitive affordances to support human memory limitations
61	A	1	The difference between the extraction options "day of the month" and "day of the week" are not clear	Translation -> Content and Meaning -> Completeness and sufficiency of meaning
62	A	1	Entering a formula to extract the day, the participant struggled to plan a formula	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
63	A	1	When an operation is selected on a column that is on furthers right side of the screen that is still visible, then the preview is not visible due to the sidebar popup blocking the original column and the preview opening outside of the visible area.	Assessment -> Issues about feedback -> Presentation -> Layout and Grouping
64	A	1	User tries to enter a single slash without quotations as a delimiter value for splitting	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
65	A	1	Feedback not aiding error recovery after entering forward-slash as a delimiter value. The feedback is a suggestion that does not fit the user's intention and does not provide smart ideas	Translation -> Existence -> Existence of a cognitive affordance-> Existence of a cognitive affordance to show the way.
66	B	1	Entering a groupby formula to filter for type upsell (syntax)	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system
67	C	1	Trying to first sort them and then count the number of leads in two seperate steps. The system mainly supports doing everything in one step with a groupby, formula or smart suggestion	Translation>Existence>Existence of a way> Missing feature
68	C	1	Writing of code in a pragmatic manner that involves table values as references rather than column variables. This way users type in their code close to what the task description in their head is 'count number of websites' I wanna type in like website and then it gives me the amount	Planning -> Users knowledge of system state,modalities -> Matching user's conception of the system

Fig. A.15.: The third part of the usability problem inventory which is based on analyzing the observation and verbalization recordings