Text

Description automatically generated with medium confidence

MSc. in Computer Science – Team Project

User Evaluation Report

Team 1 – Data Polish

Oluwatobi Omole D22125039

Sean McCrossan D22124413

Umama Sumlin Tasnuva D22124465

Naveen Maheswaran D22124491

Nikodem Adamski C18415776

Table of Contents

[1. Introduction 3](#_Toc150957064)

[2. Proposed Hypothesis 3](#_Toc150957065)

[2.1. Phase 0: Market Research and Market Analysis 3](#_Toc150957066)

[2.1.1 Market Research 3](#_Toc150957067)

[2.1.2 Market Analysis 3](#_Toc150957068)

[2.2. Phase 1: Feature Prototyping and Design of Application 3](#_Toc150957069)

[2.3. Phase 2: End-to-End Usability Testing 3](#_Toc150957070)

[3. Overview, Experimental Method, and Practical Setup 3](#_Toc150957071)

[3.1. Phase 0: Market Research and Market Analysis Data Collection 3](#_Toc150957072)

[3.2. Phase 1: Feature Prototyping and Design of Application (Interview) 3](#_Toc150957073)

[3.3. Phase 2: End-to-End Usability Testing 3](#_Toc150957074)

[3.3.1. Part 1- Cognitive Walkthrough 4](#_Toc150957075)

[3.3.2. Part 2- Think Aloud Method 4](#_Toc150957076)

[3.3.3. Questionnaire (Quantitative & Qualitative) 4](#_Toc150957077)

[3.3.4. Expert Feedback 4](#_Toc150957078)

[4. Selected Subjects 4](#_Toc150957079)

[5. Data Collection 4](#_Toc150957080)

[5.1. Phase 0: Market Research and Market Analysis Data Collection 4](#_Toc150957081)

[5.2. Phase 1: Feature Prototyping and Design of Application (Interview) 4](#_Toc150957082)

[5.3. Phase 2: End-to-End Usability Testing 4](#_Toc150957083)

[5.3.1. Part 1- Cognitive Walkthrough 4](#_Toc150957084)

[5.3.2. Part 2- Think Aloud Method 4](#_Toc150957085)

[5.3.3. Questionnaire (Quantitative & Qualitative) 4](#_Toc150957086)

[5.3.4. Expert Feedback 4](#_Toc150957087)

[6. Data Analysis and Results 4](#_Toc150957088)

[6.1. Phase 0: Market Research and Market Analysis Data Collection 4](#_Toc150957089)

[6.2. Phase 1: Feature Prototyping and Design of Application (Interview) 4](#_Toc150957090)

[6.3. Phase 2: End-to-End Usability Testing 4](#_Toc150957091)

[6.3.1. Part 1- Cognitive Walkthrough 4](#_Toc150957092)

[6.3.2. Part 2- Think Aloud Method 4](#_Toc150957093)

[6.3.3. Questionnaire (Quantitative & Qualitative) 4](#_Toc150957094)

[6.3.4. Expert Feedback 4](#_Toc150957095)

[6.3.5. AB Test - Version 1 and Version 2 4](#_Toc150957096)

[7. Conclusion 4](#_Toc150957097)

# Introduction

The report aims to define the user experience of our target market in the context of their evolving changing preferences and requirements. This report is the culmination of extensive market research and analysis, conducted with a sharp focus on understanding and meeting the intricate needs of our users and aligning with the latest academic insights in data management. Our approach was two-fold: First, we engaged directly with the users of competing data cleaning tools through a detailed survey. This survey was meticulously designed to extract not just surface-level preferences but to delve into the deeper needs, challenges, and expectations of our intended users. It provided us with rich insights, revealing areas where our tool could improve. Second, we conducted a rigorous analysis of academic literature and market trends. This step was crucial in identifying the technological frontiers our tool needed to conquer, explicitly focusing on enhancing complexity, usability, and the overall user experience. The insights gained from these endeavours informed the next critical phase of our project - feature prototyping and application design. Led by one of our team members, Umama, we embarked on creating a prototype that was functionally robust but also intuitive and user-friendly. To ensure that our prototype resonated with our intended audience, we conducted semi-structured interviews with data science students, employing the methodologies outlined in 'The Mom Test'. This approach was pivotal in minimizing biases and maximizing the value of user feedback. The interviews yielded a plethora of information, highlighting both the strengths and areas for improvement in our prototype - from the functionality of our data cleaning features to the versatility in file format support and the depth of data profiling capabilities. This report presents a detailed account of our journey from conceptualization to prototype, the invaluable feedback we received, and the strategic enhancements we plan to implement. It is a testament to our commitment to delivering a data cleaning tool that is not only state of the art but also profoundly attuned to the needs of its users.

# Proposed Hypothesis

## Phase 0: Market Research and Market Analysis

### 2.1.1 Market Research

To better understand our end users' needs and desires and academic advances. Two methods of market research were evaluated:

1. A survey was sent out to the end users of competitor data cleaning tools used in their day-to-day activity with data cleaning. This survey was composed of questions that allowed us to measure our end user and questions that better helped us understand if the end-user was already equipped with using software to clean data.
2. The second step was to analyze the market in the academic sphere. In this stage, four papers were identified. From those papers, we realized that our tool requires more complexity, usability and UX improvements to keep up with the ever so changing technology.

### 2.1.2 Market Analysis

We first decided to take the best results from the Google search engine during the market analysis. This initial set of competitors then evolved into receiving expert feedback from experts in the field of data cleaning on the data cleaning tools we have decided to choose. Here is the competitor matrix that stemmed from our market research, as seen in [Figure 1].

A screenshot of a chart

Description automatically generated

Figure 1 Competitor matrix (post expert feedback)

## Phase 1: Feature Prototyping and Design of Application

The prototype created by Umama, one of our team members, was then formulated into interview-style questions where semi-structured interviews were held with four individuals who met our user profile, data science students.

The survey was created using the famous novel 'The Mom Test'. This novel was made for entrepreneurs to get the most out of any end-user interaction that minimizes bias and maximizes information gain.

The interview results were initial positive feedback on our application journey, name, application design and application features. Initial negative feedback on our data cleaning function as, at the time, we had one-button cleans all. Users wanted the ability to clean individually, the import-export being limited to CSV and not extending to Excel. Our data profiling page did not contain any statistical insight optional for the user who would like to dive deep into their dataset.

These comments were considered when determining the appropriate application enhancements that satisfy the end users and academia.

## Phase 2: End-to-End Usability Testing

# Overview, Experimental Method, and Practical Setup

## Phase 0: Market Research and Market Analysis Data Collection

Overall, what was received from phase 0 is understanding our end users, if there is demand for our product, and ranking the top features that end users expect from that product.

## Phase 1: Feature Prototyping and Design of application (Interview)

Overall, what was received from phase 1 is a more in-depth understanding of our UI, UX and features list that satisfy our users. This feedback was crucial in determining the quality of our application and a solidifying factor that we are indeed on track to achieving our desired outcome of success. As mentioned before, the interviews were semi-structured, which means that I sent out a list of questions initially to our four candidates; they were prepared, but I did not show our prototype, so when interviews came, their answers came naturally, and more conversational approach was made when users, for example, asked me questions or moved away from the topic we still noted it down. We did not interrupt their talking and strain of thought.

## Phase 2: End-to-End Usability Testing

### Part 1- Cognitive Walkthrough

For the cognitive walkthrough, the users were given a set of tasks to complete. Each user was given the same tasks. Each of the selected users fits into the target user group. The cognitive walkthroughs were conducted in person and in a quiet place using the user's laptop. The product was explained to each of the users before starting.

How long it took the user to complete the tasks, notes on how the user completed the task, whether they failed and what feedback the user had was recorded on a spreadsheet. The users completed two sessions, each one after the other, for the two versions of the product.

The decrease in time to complete each task in the second version can be partly attributed to the user becoming familiar with the tasks and that future cognitive walk-throughs should use different users.

The tasks to complete were as follows:

Cognitive walkthrough tasks Version 1 (Job Posting)

* How would you upload the data?
* How would you check which column has the most missing values?
* How would you check the outliers in your numeric data?
* How would you determine the quality of your data?
* How would you clean the data (remove duplicates)?
* How would you export the new clean data?

Cognitive walkthrough version 2 (New UI)

* How would you upload the data?
* How would you check which column has the most missing values?
* How would you check the outliers in your numeric data?
* How would you determine the quality of your data?
* How would you clean up the missing values?
* How would you clean the rest of the data based on the profiling you saw on the profiling page?
* How would you export the new clean data?

### Part 2- Think Aloud Method

The think-aloud method is a user evaluation technique where individuals are asked to verbalize their thoughts as they go through a specific task. The think-aloud way is highly beneficial for obtaining user evaluations of our web application as it allows us to gain direct insights into users' cognitive processes and decision-making. By encouraging users to express their thoughts while interacting with the application, we can identify usability issues in real time, understand their feelings and satisfaction levels, and validate design choices. This approach is cheap and easy to conduct and will help us create a more user-friendly and effective web application.

For both versions of Data Polish, think-aloud evaluation was carried out with two users per version. These user evaluations were carried out online as it was more flexible for the user to attend them at convenient times. Users were provided with a brief description of the application's functionalities and instructions at the beginning of each session. Throughout the process, team members noted down the users’ perspectives.

Along with speaking about the users' thoughts, for each of the pages, they were asked mainly to answer the following four questions:

1. What aspects do you LIKE about this page?
2. What aspects do you DISLIKE of this page?
3. What CHANGES would you suggest for this page?
4. Do you have any ADDITIONAL comments for this page?

### Part 3- Questionnaire (Quantitative & Qualitative)

Qualitative and quantitative surveys were created using the famous novel on questionnaires called 'A.N. Oppenheim, Questionnaire Design.'. the feedback and changes made from the book were two separate two questionnaires; this is a crucial part as it allows us to focus on whether our data lack consistency, e.g. if a user would quickly answer 4/10 in a quantitative survey. However, in a qualitative one, they would praise the overall performance of the UI, meaning the actual answer should have been closer to 7 or 10, but it was not; this gap could then get us to understand better why this inconsistency exists and what could we do with it.

### Part 4- Expert Feedback

The process implemented for the expert reviews was structured around Nielsen’s usability heuristics, which represent fundamental principles for user interface design. The selected expert reviewers, equipped with their own devices, were allowed to conduct their evaluations at a time and place.

Each reviewer had a custom-designed evaluation table to facilitate a detailed and systematic analysis. This table was divided into two primary columns to ease the assessment process. The first column was specifically allocated for the severity rating of each heuristic violation encountered. This rating system was intended to quantify the impact of each issue on the overall user experience, ranging from minor concerns to critical usability flaws.

The second column of the table was dedicated to collecting comprehensive comments and observations from the reviewers about each problem they identified. This column was designed to capture in-depth qualitative feedback, providing insights into the nature of the issues, the context in which they occurred, and potential implications for users.

# Selected Subjects

The selected subjects for our data cleaning application came from the initial survey feedback. We have decided to stay with three types of subjects: Data science students and Data professionals. The selected subjects were then followed up for future interviews, cognitive walkthroughs, ab testing and finally think alound protocals along side surveys. When choosing subjects we decided to only choose between these two subject. Gender is non-important, age is non-important, data knowledge is important, and data usage is crucial. We wanted individuals who are already aware of data cleaning stress and pressure for them to achieve relief.

# Data Collection

## Phase 0: Market Research and Market Analysis Data Collection

The data collection was in the form of the initial survey, which allowed us better to understand our target audience, their age, gender, what features they would like to see and their rank. The initial survey was both qualitative and quantitative, providing us with an excellent heads-up advantage overview of our target audience.

## Phase 1: Feature Prototyping and Design of application (Interview)

The data collected from our end users during the interviews were feedback on our application design, name, individual page feedback, and journey. They added new features such as a dummy categorical feature cleaning statistical insight feature to see the normal distribution and overall understanding of what is expected from the user journey through our application.

## Phase 2: End-to-End Usability Testing

### Part 1- Cognitive Walkthrough

For the cognitive walkthrough evaluation, qualitative and quantitative data were collected to give insight into how users interacted with the product and how long it took them to complete core tasks within the product. This data helps us determine whether our product is easy to use and quicker than traditional coding techniques for cleaning data. The users were then asked for feedback after completing these core tasks.

Gathering this data is important to answer key questions such as “Is our product more efficient than traditional coding techniques?” "Does our product give the users a better understanding of their data quality issues?". These insights will allow us to improve our product and give the user a better experience.

### Part 2- Think Aloud Method

For think-aloud evaluation, qualitative data were collected, which gives detailed insight into users’ thoughts about each page or feature of the application. This data helped us to understand how users felt and what they liked or found challenging. We also asked them to suggest any changes or improvements to make the features more user-friendly.

During the design and development of our application, as creators, we can sometimes see it from a specific perspective. Gathering input from various users offers fresh insights and highlights aspects we may have yet to consider. What seems straightforward might be different for someone using the application for the first time. Exploring diverse user perspectives helps us uncover potential improvements and ensures our application is user-friendly and accessible to a broad audience.

### Part 3- Questionnaire (Quantitative & Qualitative)

For questionnaires, we received valuable insight on our UI, UX and features list, page by page. This feedback is an essential factor in better our overall product, and our being so user-driven has enhanced our ability to deal with such feedback. The data collected followed a rating system of our interface, a features list for data cleaning and profiling, and a better understanding of the user journey and whether we fit their satisfaction.

### Part 4- Expert Feedback

1. **Quantitative Data Collection**

The expert evaluators gathered quantitative data. They were tasked with quantifying each heuristic violation identified during their application review. This quantification was executed using a clearly defined scale that ranged from 0 to 4. On this scale, a rating of '0' signified no violation of the heuristic principles, indicating that the aspect under review was in full compliance with Nielsen's usability guidelines. Conversely, a rating of '4' represented a severe breach of usability norms, categorized as a 'usability catastrophe.' This severity rating aimed to provide a measurable, objective assessment of each identified issue, allowing for a systematic evaluation of the application's overall usability.

1. **Qualitative Data Collection**

Qualitative data was also collected during this stage. This data stemmed from the detailed comments and observations provided by the evaluators. Accompanying their scaled evaluations, these comments offered rich insights into each heuristic violation. They accounted for the evaluators' in-depth analysis, personal observations, and professional judgments regarding the nature and implications of the issues encountered. This qualitative feedback was essential in adding context and depth to the numerical ratings.

# Data Analysis and Results

## Phase 0: Market Research and Market Analysis Data Collection

The analysis done on the market research analysis was understanding key features found within our competitor applications. We also identified the overall cleaning structure and how to approach cleaning itself. Finally, we also had to look through journal articles to see if any key features are mentioned in journal articles that suggest some to be deemed essential and others to be extra. Through surveys, we could understand key features quickly; this was important as it saved us much time and verified what other applications and journal articles already had; with this clarification, we could be confident and continue forward.

## Phase 1: Feature Prototyping and Design of application (Interview)

The interview analysis breaks down words using TF\_IDF. We had to dive deep into the comments they used and the suggestions they made. The love and hatred they had for a particular page, idea, or feature were all essential in understanding what our end users wanted and why they wanted it. Through this, we could quickly pivot on our ideas, such as static images, one button cleaning all and adding colour to our application. Key steps that would only have been done if told by our end user.

## Phase 2: End-to-End Usability Testing

### Part 1- Cognitive Walkthrough

Analyzing the qualitative and quantitative data collected during the cognitive walkthrough is crucial to understanding how users interact with the product. Participants were asked to complete specific tasks in the product. Time was recorded to know how long it took them, notes were recorded to document whether they had difficulty or required intervention after each task, and feedback was collected.

#### 6.3.1.1. Version 1

The following results are from the cognitive walk-throughs using version 1 of the product.

|  |  |  |
| --- | --- | --- |
| Task 1 - Upload the data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 38 | 95 |
| Notes from the task: | The website did not work on my laptop due to security screening. The user initially tried to upload an Excel file but got an error. I had to intervene and tell them only comma-delimited CSVs are accepted. The user also did not understand Job ID and had to navigate back to retrieve it. | The user tried to upload Excel, but intervention was required to explain only comma-delimited CSV's and had to tell the user about the Job Posting ID. |
| Feedback from user: | Should accept Excel. Does not like Job Posting ID. | Did not like the Side Bar to navigate. Not liking the Job Posting ID. Overall felt clunky |

|  |  |  |
| --- | --- | --- |
| Task 2 - Which column has the most missing values? | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 394 | 112 |
| Notes from the task: | Eventually found the data profile page for the task. There was no continue button, so I had to navigate manually. I had trouble checking the status and the job posting ID before seeing the results—required intervention. I found the correct column. | Check the status first; the user is frustrated. Only able to do it with intervention. |
| Feedback from user: | Very slow, needed to understand the Job ID, UI needed to be more intuitive after uploading the file to see the quality of their data. The visual was okay and did the job. | Casing is wrong. I could hardly read the font. The check status needed to be clearer, and the job posting ID. |

|  |  |  |
| --- | --- | --- |
| Task 3 - Check the outliers in your data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 11 | 15 |
| Notes from the task: | Able to quickly find the outliers as already on the profiling page. | Able to quickly find the outliers |
| Feedback from user: | It took a lot of work to determine what outliers were and how many they had. | Does not like the chart. |

|  |  |  |
| --- | --- | --- |
| Task 4 - Determine the quality of your data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 273 | 42 |
| Notes from the task: | They struggled to determine their data quality and saw they had many missing values and high cardinality in fields that they expected but needed help to give a confident answer. Fail | I need help to make a confident decision. |
| Feedback from user: | More information is needed to determine if the quality of data is good. There should be a summary, and all visuals should be on the page so that you can make an informed decision. | I need to find out the quality of the data. Visuals are hidden and hard to navigate and remember. |

|  |  |  |
| --- | --- | --- |
| Task 5 - Clean your data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 55 | 73 |
| Notes from the task: | I was able to navigate more easily but had to go back and get the job posting ID after the intervention | I had to be told about the job posting ID. The user needed to learn that the data was cleaned. |
| Feedback from user: | Job ID posting is confusing but is easy to use. | The error message is all in lowercase, which is annoying. Does not understand the Job ID and needs to be faster. It doesn't give any feedback when cleaned. |

|  |  |  |
| --- | --- | --- |
| Task 6 - Export the data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 17 | 8 |
| Notes from the task: | The job posting ID was confusing, but the user was able to export data with no help | The job posting ID was confusing, but the user was able to export data with no help |
| Feedback from user: | Job posting ID is confusing again but easy to use | The export button is not aligned like the other |

#### 6.3.1.2. Version 2

The following tables are the results from the cognitive walk-throughs using version 2 of the product.

|  |  |  |
| --- | --- | --- |
| Task 1 - Upload the data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 13 | 7 |
| Notes from the task: | No issues | No issues |
| Feedback from user: | They need to know what they're allowed or not allowed to upload. | It looks smooth and better than the previous one. |

|  |  |  |
| --- | --- | --- |
| Task 2 - Check missing values | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 14 | 15 |
| Notes from the task: | I found the column with no help | I went through profiling, but there was a significant delay. Eventually, it came up for profiling |
| Feedback from user: | Easier than the first one. | A weird error message says it was profiled, but then I went to check an error. |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Task 3 - Check outliers | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 8 | 5 |
| Notes from the task: | I found the columns with outliers with no help | Found the columns |
| Feedback from user: | I have the same issue with the visuals, they should be clearer. | Visuals are okay |

|  |  |  |
| --- | --- | --- |
| Task 4 - Determine data quality | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 187 | 43 |
| Notes from the task: | Same as previous | Same as previous |
| Feedback from user: | (SAME) There is not enough information to determine if the data quality is bad. There should be a summary, and all visuals should be on the page so that you can make an informed decision. | Unable to determine the quality |

|  |  |  |
| --- | --- | --- |
| Task 5 - Clean the missing values | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | - | - |
| Notes from the task: | - | - |
| Feedback from user: | - | - |
|  |  |  |
| Task 6 - Clean the data based on the results you saw | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | - | - |
| Notes from the task: | - | - |
| Feedback from user: | - | - |

|  |  |  |
| --- | --- | --- |
| Task 7 - Export the data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 9 | 5 |
| Notes from the task: | Exported the data | Exported the data |
| Feedback from user: | Export worked fine | Export worked fine |

### Part 2- Think Aloud Method

Analyzing the qualitative data collected from users during the think-aloud evaluation was a crucial step in understanding the user experience with our application. We carefully reviewed participants' verbal expressions, comments, and feedback to extract meaningful patterns and insights. We have categorized the data for two versions separately. For each feature or page, we categorized liking, disliking, suggested changes, and additional comments of user one and user 2.

#### 6.3.1.1. Version 1

The tables below outline the user's categorised data for think-aloud evaluation of version 1 of the application.

|  |  |  |
| --- | --- | --- |
| **Step 1: Data Upload** | | |
|  | User 1 | User 2 |
| Like | Drag & drop – file upload was faster as the folder was already opened. | Simple and conventional layout |
| Dislike | * Only .csv is supported (the user had a .xlsx file, so they had to convert the file before uploading) * I just remembered that I need to copy the job ID. I had to go to the Home page to copy the job ID * I was expecting it to automatically go to the next page after uploading the file | Hates copying job ID (the user feels it wastes time. Also if some other text is in between the usage of this application, then that text will be in the clipboard, so the user needs to copy job ID again) |
| Changes Suggested | Better home page like a "Welcome" page instead of a "Job ID" page | Better to have a button to go to the next page instead of choosing from the sidebar. |
| Additional comment | none | Overall, it is easy to use except for having to copy the job ID |

|  |  |  |
| --- | --- | --- |
| **Step 2: Data Profile** | | |
|  | User 1 | User 2 |
| Like | * Useful charts * While hovering, the red border distinguishes the selected part and helps focus. | Option to download the image, so no need to take a Screenshot |
| Dislike | * Colour of the "Check Status" button * I expected it to go to the next page after uploading the file. | * Pasting job ID to check the status * No button to go to the next page |
| Changes Suggested | The user was expecting to view the dataset after uploading before proceeding further. So, a page showing the uploaded dataset would be helpful. | Instead of clicking and opening each image individually, it will save the number of clicks if all the images are shown on the page. |
| Additional comment | It is good to get some insights about the condition of data before proceeding to clean. | Look-wise, it didn't feel exciting. More colours would be nice. |

|  |  |  |
| --- | --- | --- |
| **Step 3: Data Cleaning** | | |
|  | User 1 | User 2 |
| Like | none | none |
| Dislike | * Tired of pasting job ID * The page looks too empty | * Again Job ID * Fundamental cleaning as it is only removing duplicates * No button to go to the next page |
| Changes Suggested | As a data-cleaning tool, there should be a lot of options on this page | There should be a button for the next export page or text saying the user is supposed to go to export after this page. As the button says "Start Data Cleaning", it is confusing what is expected to be done after starting the process |
| Additional comment | none | none |

|  |  |  |
| --- | --- | --- |
| **Step 4: Data Download** | | |
|  | User 1 | User 2 |
| Like | Downloading file properly as expected | Download is working fine |
| Dislike | * Too often, I had to check the process completion status by pasting the job ID. * Why "Check status" button is red? It felt as if it was dangerous to click on the button to check the status (was said in a joking manner) | * Again Job ID |
| Changes Suggested | There can be an option/choice for the user to download a file in either .csv or .xlsx format. | All is good. Remove the existence of the job ID entirely from the application |
| Additional comment | The concept and plans of the entire application seems to be very easy and helpful for data-cleaning purpose | none |

#### 6.3.1.2. Version 2

The tables below outline the categorized data from users for think-aloud evaluation of version 2 of the application.

|  |  |  |
| --- | --- | --- |
| **Step 1: Data Upload** | | |
|  | User 1 | User 2 |
| Like | * It is good to see the progress bar while the file is being uploaded * Easy to navigate to the next page | The arrows make all the steps from Import to Export easy to follow. The arrow also highlights on which page the user is currently |
| Dislike | Only .csv is supported | Only .csv is supported |
| Changes Suggested | * The logo and application name should be visible * The previous colour theme was more excellent than the current black one | none |
| Additional comment | none | none |

|  |  |  |
| --- | --- | --- |
| **Step 2: Data Preview** | | |
|  | User 1 | User 2 |
| Like | * So many features on one page! | * Loads of valuable actions can be performed along with previewing data * A good amount of rows are shown * The file can be exported after making some changes in CSV or Excel format |
| Dislike | none | none |
| Changes Suggested | none | none |
| Additional comment | This page has a lot more than it was expected to have | Excellent |

|  |  |  |
| --- | --- | --- |
| **Step 3: Data Profile** | | |
|  | User 1 | User 2 |
| Like | * Having the option to return to the data preview page from here * Data quality metric * Interactivity of charts * Colourful and interesting | * Captivating way of providing data analysis * Enough number of charts |
| Dislike | It might be too noisy when the colour theme is changed | Nothing if it is completed as per plan |
| Changes Suggested | none | none |
| Additional comment | Quite detailed data profiling is shown | none |

|  |  |  |
| --- | --- | --- |
| **Step 4: Data Cleaning** | | |
|  | User 1 | User 2 |
| Like | All essential cleaning options, including the advanced cleaning features, are present. | * Various data cleaning options are available * The data preview section at the bottom * Adjustable panels |
| Dislike | * With a more significant number of rows, you have to scroll sideways a lot * The "Date" data type is not supported, though it is a prevalent one * No automatic cleaning feature | * No automatic cleaning feature * Too many clicks can be time-consuming, even if the user wants some simple cleanup. |
| Changes Suggested | Instead of scrolling sideways, scrolling down would be easier, just like it was in the prototype | Adding some one-click button for simple automatic cleaning would be good to have |
| Additional comment | none | none |

|  |  |  |
| --- | --- | --- |
| **Step 5: Data Download** | | |
|  | User 1 | User 2 |
| Like | Perfectly working | Download is working fine |
| Dislike | * No option to download in xlsx format | * Colour of the “Export as CSV” button * Downloaded only in CSV format, whereas data preview had both CSV and Excel formats |
| Changes Suggested | * A progress bar to show the progress of data cleaning would be better to indicate when data is ready for downloading. * There can be an option/choice for the user to download the file in either .csv or .xlsx format, just like the data preview page   . | * After finishing all the steps, a happy, successful icon or emoji would be nice. * Add a login system to store the history of datasets worked with |
| Additional comment | I enjoyed the experience of using the entire application | After necessary changes are made, data science students' lives will be easier. Good application. |

### Part 3- Questionnaire (Quantitative & Qualitative)

Here are the results of the qualitative and quantitative surveys:

Quantitative survey results are as follows:

Forms response chart. Question title: Please select your current role:
. Number of responses: 6 responses.

Forms response chart. Question title: Is this your first time using a data-cleaning application?
. Number of responses: 6 responses. Forms response chart. Question title: How would you rate the overall design of the application?
. Number of responses: 6 responses. Forms response chart. Question title: Rate the user-friendliness of the following pages.
. Number of responses: . Forms response chart. Question title: Rate the user interface&apos;s ease of navigation.
. Number of responses: 6 responses. Forms response chart. Question title: How would you rate the usefulness of the following Data Cleaning Features?
. Number of responses: . Forms response chart. Question title: How effective are the Outlier Management features in your data cleaning process?
. Number of responses: . Forms response chart. Question title: Rate the ease of exporting data to CSV format.
. Number of responses: 6 responses. Forms response chart. Question title: Overall, how satisfied are you with the Data Polish application?
. Number of responses: 6 responses. Forms response chart. Question title: On a scale from 1 to 5, how likely are you to recommend the Data Polish application to a friend or colleague?
. Number of responses: 6 responses.

On the qualitative survey, we failed to get anyone to answer due to the length of each question, so if any question was deemed necessary, we asked them verbally instead of written.A screenshot of a phone

Description automatically generated

### Part 4- Expert Feedback

The expert reviewers for this section consist of the three supervisors for the Master's Group project option. Over the course of weekly meetings and designated show-and-tell presentations, they have given feedback, which has been fed back into product development as an iterative loop. Their feedback has been collated below and tagged under their relevant headings.

1. **Frameworks & Data Quality**:
   * DAMA and Zachman's frameworks highlight structured methodologies for ensuring data quality.
   * Emphasizes the role of governance and standardized processes in maintaining data integrity.
2. **User Personas & Needs**:
   * Survey and interview-derived user personas guide tool customization, ensuring they meet varied user requirements.
   * Highlights the importance of understanding user demographics for tool effectiveness and selecting a target user.
3. **Competitor Research and Insights**:
   * Comparative analysis of tools like Open Refine and Oracle Analytics underscores the necessity for adaptable, robust data-cleaning solutions.
   * Indicates a preference for tools with broad compatibility and comprehensive data handling capabilities like the tools above\
   * Studying the user experience of these tools highlights common themes in developing this type of app to adopt and other issues to work on.
4. **Robust Data Importation Strategies**:
   * Focuses on the capacity to handle unstructured data and various file formats.
   * Stresses the need for flexible import features to accommodate diverse data types.
5. **Advanced-Data Cleaning Techniques**:
   * Implement outlier detection, profiling, and visualization techniques for thorough data cleaning.
   * Suggests leveraging these advanced methods for deeper data analysis and cleaner datasets.
6. **User Interface Feedback**:
   * Highlights include the need for simplicity, intuitive design, and easy navigation to enhance user experience.
   * Users expressed a preference for interfaces that provide clear visualizations of data and straightforward access to key features.
   * There's noted importance of customizable dashboards and the ability to handle complex tasks with minimal clicks or navigation.
   * Feedback suggests that responsive and adaptive interfaces significantly improve user engagement and productivity.

### 6.3.5. A/B Test - Version 1 and Version 2

In conducting an A/B test between version 1 and version 2 of our application, we systematically employed both think-aloud evaluations and cognitive walkthroughs for each performance. This comprehensive approach aimed to capture user perspectives through verbalized thoughts and assess the mental aspects of interaction. Through the think-aloud evaluations, we gathered qualitative insights, allowing users to express their experiences, preferences, and concerns. Simultaneously, the cognitive walkthroughs provided a structured analysis of the application's usability by simulating user interactions. This section summarises our combined findings using both evaluation methods for each version respectively.

#### 6.3.5.1 Version 1 - Summary

Overall, the users did not enjoy using version 1 of our application. The following are the main points that came up throughout each of the tasks:

1. Job Posting ID – The users hated putting in their Job Posting ID before doing any tasks, which interrupted their workflow. The interviewer had to intervene numerous times as the user was initially unable to figure it out.
2. UX – Job Posting ID was the main culprit, but the users did not like how it flowed and would have preferred to have had a continue button rather than navigating through the sidebar.
3. Data profiling – The user felt the charts are too basic and do not give enough information to determine the quality of their data. They found it hard to remember all their issues and had to navigate through each of the charts numerous times to conclude whether the quality of their data was poor and needed fixing.
4. UI inconsistency – The alignment of buttons and text casing was noticed, making the product look unprofessional.
5. File import – Both users initially tried to upload an Excel file. There was no message to say what files were accepted.

#### 6.3.5.2 Version 2 - Summary

Straight after the first interview, the same users were shown the second version, where the Job Posting ID and the sidebar were removed. However, the data profiling plots were the same.

Overall, the users found version 2 a lot more usable, but they still had some concerns that are listed below:

1. Buggy and slow – The Data preview page took a long time to load up and had to go back and forth between the pages before seeing their data previewed.
2. Data cleaning – This version is highly buggy and didn't work, and the interview subjects didn't want to do it anymore.
3. Determine the data quality – This was not improved from the first version, and users had the same complaints: charts are hidden and not thorough enough to determine whether the quality of their data is poor.

#### 6.3.5.3 Comparison between version 1 and version 2

Decrease in time taken to complete tasks - When comparing the times for versions 1 and 2, we can see that users completed the tasks much quicker, especially when navigating the app.

Some of these changes can be attributed to completing the second version straight after the first, but most were through UX improvements and the Job posting ID removal.

Data profile – The data profiling and determining the data quality is poor in both versions. Neither subject was confident in their answer. Steps are being taken to improve this. The next version will have more charts and be on the same page. We are also looking at creating a data quality metric which the user can reference when determining the quality of their data.

UI – UX got a lot better feedback in version 2, but users complained that all black for the UI made it look almost dangerous and had viruses…

# Conclusion

In conclusion, the journey detailed in the report culminates in a profound understanding of the current landscape and future direction for our data cleaning tool. This comprehensive exploration, rooted in a blend of market research and academic analysis, has provided us with a clear roadmap for development. The insights derived from our user survey have been remarkably revealing, offering us a view into our tool's real-world applications and expectations. Users have voiced their need for more nuanced data cleaning functionalities, highlighting the importance of individualized cleaning options over a one-size-fits-all approach. Furthermore, the feedback emphasized the need for broader file format compatibility, extending beyond CSV to include formats like Excel, catering to a broader range of user preferences and requirements. Equally important was a more robust data profiling feature, allowing users to delve deeper into their datasets with comprehensive statistical insights. These insights are not just isolated observations but are integral to our vision of creating a tool that is not just functionally superior but also more user-centric to our competitors. Our commitment to this vision is reflected in the strategic enhancements we plan to implement, ensuring that our tool is not only in step with current demands but also anticipates future trends in data science. The findings and strategies outlined in this report will serve as a roadmap as we move forward. They will inform our developmental efforts, ensuring we continue innovating and elevating the user experience.