# Cognitive walkthrough

Two subjects were interviewed using the cognitive walk-through technique. Each user was given several tasks to complete. Time, notes and feedback was collected for each of these tasks. It is important to note that this experiment is imperfect as for the second version the users knew what they had to do and were a lot quicker. They did use separate datasets, however.

## User 1 profile

User 1 is a lead data analyst for a major airline and a part time data science student. They primarily work with excel files and must prepare analysis for different departments in his company.

## User 2 profile

User 2 is a pharmacist who is pursuing a master’s in clinical information systems studies. They currently have low coding ability but as part of their course are required to learn and analyze data in R. This user also has dyslexia.

## Version 1

### Summary

Overall, the users did not enjoy using version 1. A few main points came up throughout each of the tasks

1. Job Posting ID – The users hated having to put in Job Posting ID before doing any tasks and it really 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, would have preferred to have had a continue button, rather than navigating through the side bar.
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 bad and needed fixing.
4. UI inconsistency – Alignment of buttons and text casing was noticed and made the product look unprofessional.
5. File import – Both users initially tried to upload an excel file. There was no message to say what files are accepted.

|  |  |  |
| --- | --- | --- |
| Task 1 - Upload the data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 38 | 95 |
| Notes from task: | Website did not work on work laptop due to security screening. User initially tried to upload an excel file but got an error. Had to intervene and tell them only comma delimited CSV's are accepted. User also did not understand Job ID and had to navigate back to retrieve it. | User tried to upload excel but intervention required to explain only comma delimited CSV's. Had to intervene and tell 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. Did not like 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 task: | Eventually found data profile page for the task, there was no continue button, had to manually navigate. Had trouble with checking the status and the job posting ID before seeing the results. Required intervention. Found the correct column. | Check status first user is frustrated. Unable to do it without intervention. |
| Feedback from user: | Very slow, did not understand the Job ID, UI was not intuitive after uploading file to see the quality of their data. Visual was ok and did the job. | Casing is wrong. Could hardly read the font. Check status was very confusing and did not know what job posting ID was. |

|  |  |  |
| --- | --- | --- |
| Task 3 - Check the outliers in your data | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 11 | 15 |
| Notes from task: | Able to quickly find the outliers as already on profiling page. | Able to quickly find the outliers |
| Feedback from user: | Visuals were hard to determine what were outliers 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 task: | Struggled to determine their data quality said they saw they had a lot of missing values and high cardinality in fields that they expected but could not give a confident answer. Fail | Is not able to make a confident decision. |
| Feedback from user: | Not enough information to determine if quality of data is bad. Should be a summary and all visuals should be on the page so that you can make an informed decision. | Cannot tell the quality of 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 task: | Was able to navigate easier but had to go back and get job posting ID after intervention | Had to be told about the job posting ID. User did not know that data was cleaned. |
| Feedback from user: | Job ID posting is confusing but is easy to use. | Error message is all in lower case which is annoying. Does not understand the Job ID and is too slow. Doesn’t give any feedback when cleaned. |

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

## Version 2

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

### Summary

Overall, the users found version 2 a lot more usable, but they still had some concerns

1. Buggy and slow – Data preview page took a long time to load up and had to go back and forth the pages before seeing their data previewed. However, this wasn’t a task so not recorded but when interviewing this was the biggest issue.
2. Data cleaning – This version is extremely 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 bad or not.

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

|  |  |  |
| --- | --- | --- |
| Task 2 - Check missing values | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 14 | 15 |
| Notes from task: | Found the column with no help | Went through profiling but was a significant delay. Eventually came up for profiling |
| Feedback from user: | Easier than the first one. | Weird error message say it was profiled but then went to check and error |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Task 3 - Check outliers | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 8 | 5 |
| Notes from task: | Found the columns with outlier with no help | Found the columns |
| Feedback from user: | Same issue with the visuals, thinks they should be clearer | Visuals are ok |

|  |  |  |
| --- | --- | --- |
| Task 4 - Determine data quality | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | 187 | 43 |
| Notes from task: | Same as previous | Same as previous |
| Feedback from user: | (SAME) Not enough information to determine if quality of data is bad. 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 4 - Clean the missing values | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | Bug | Bug |
| Notes from task: | Bug | Bug |
| Feedback from user: | Bug | Bug |
|  |  |  |
| Task 4 - Clean the data based on results you saw | | |
| User: | User 1 | User 2 |
| Time taken (seconds) | Bug | Bug |
| Notes from task: | Bug | Bug |
| Feedback from user: | Bug | Bug |

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

## Overall Summary

Decrease in time taken - When comparing the times for version 1 and 2 we can see that users completed the tasks a lot quicker, especially when navigating the app.

Some of this change can be attributed to completing the second version straight after the first but most was through the improvements in the UX and removal of the Job posting ID.

Data profile – The data profiling and determine the quality of the data 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 will all 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.

Data Cleaning – Version 2 was not stable in this section; action is being taken to fix.

Data Preview Page – Current version is too slow, and users can’t refresh without leaving the page. Either needs to be faster or have a refresh button or some sort of status loading message to notify the user when it’s done.

UI – UX is a lot better, but users complained about all black for the UI made it look almost dangerous and that it had viruses…

The users – Each of the users were in different stages of their data professional career and it showed in the interviews. User 2 was a lot less experienced and has only been learning about data analysis the past couple of months. User 1 was a lot more critical in parts of the data profiling and charts while user 2 was less so as they didn’t have the experience, but both had similar problems in both versions.