# Conclusion

Our team employs an agile with scrum methodology for the management of the project (Hu, Yuan, & Zhang, 2009). Agile with scrum was chosen as it offers flexibility, adaptability and is a more collaborative approach when compared to traditional methodologies such as waterfall. Waterfall has documented limitations including poor visibility, can cause a lack of communication between stakeholders and is bad at handling unstructured code (Gabaorv, et al, 2021).

The team is composed of two advanced software development (ASD) MSc students and three data science (DS) MSc students. ASD members work on the backend and frontend of the product while DS members work on data cleaning functionality and the data profiling dashboard. Below is a breakdown of the team roles

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Student Number | Development role | PM Roles | Testing roles |
| Sean McCrossan | [D22124413](mailto:d22124413@mytudublin.ie) | Floater between data cleaning and profiling | Developer | Tester |
| Naveen Maheswaran | D22124491 | Backend developer | Developer | Tester |
| Oluwatobi Omole | D22125039 | Data cleaning python developer | Developer | Tester |
| Nikodem Adamski | C18415776 | Data profiling python and JavaScript developer | Scrum master/Project manager | Evaluator |
| Umama Sumlin Tasnuva | D22124465 | Frontend developer | Developer | Evaluator |
|  |  |  | \*Roles rotate every sprint | |

A sprint cycle lasts one week, and a daily scrum (15-30 minute) meeting is conducted to report on progress and challenges from each team member. Jira, GitHub and Slack are used to manage the project along with requirement spreadsheets. GitHub is the project repository and is used to track pull requests and commits. Jira is used to create epics, stories and issues which are the tasks that are assigned at the sprint kick off meeting every Monday and progress is tracked throughout the week in the daily scrum. Slack is the teams internal messaging tool. GitHub, Jira and slack are all linked to each other.

A graph showing a graph

Description automatically generated with medium confidence

Figure Jira issue cumulative report

Diagram of scrum process diagram

Description automatically generated

Figure 2 Sprint cycle (Srivastava, Bhardwaj and Saraswat, 2017)

The development of the project is scheduled to stop on the 1st of December to allow sufficient time for testing, evaluation and improvements before the final release.

|  |  |  |  |
| --- | --- | --- | --- |
| Week No | Week | Features to be included | Comments |
| 2 | 25th - 29th Sep | Data Import | Start of work and research |
| 3 | 2nd - 6th Oct | Data Quality checks and export of data | Data quality check module created, not all functions made it to final |
| 4 | 9th-13th Oct | Data cleaning | Data cleaning module created, only dataset wide cleaning up taken |
| 5 | 16th - 20th Oct | Data Profiling images | D3 dashboard scrapped for this release, only static pyplots |
| 6 | 23rd - 27th Oct | UI improvements and stability | Worked on stability |
| 7 | 30th - 3rd Nov | Data Preview and Data Import V2 | Data preview to be finished |
| 8 | 6th - 10th Nov | Data profile dashboard and data cleaning column level | D3 dashboard and column level cleaning |
| 9 | 13th - 17th Nov | Create new columns, Dimension reduction | Not finalised but more complex features to be released |
| 10 | 20th - 24th Nov | Statistical analysis data profile visuals and annotation | More complex charts in data profiling |
| 11 | 27th - 1st Dec | Multiple file imports | Allow users to upload multiple files |
| 12 | 4th - 8th Dec | Testing and evaluation | Small improvements or possible new functionality (code freeze) |
| 13 | 11th - 15th Dec | Testing and evaluation | Stop all functionality (code freeze) |
| 14 | 18th - 22nd Dec | Final report writing and presentation | Product stable and general release. |

## Team meetings:

The meeting schedule follows the agile scrum methodology (Srivastava, Bhardwaj and Saraswat, 2017). Daily scrum meetings are 30 minutes that occur online every Tuesday, Wednesday and Thursday. Each team member answers the following questions:

1. What did you do yesterday?
2. What are you planning to do today?
3. Are there any blockers impacting your progress?

On these meetings team members are encouraged to share ideas or research they have completed, giving the team visibility and raise other issues in front of the team. The team has a sprint planning meeting every Monday with the objective of assigning tasks to team members to complete during the sprint. On Friday the team has two in person meeting post the lectures. On the sprint review meetings, we discuss how the current sprint went and what should the goal of the following sprint should be and then we organize the project backlog of Jira’s. On the retrospective we anonymously list, to ensure an open dialogue, what went well, what did not go well and what we should start doing, each of these points is then discussed by the team.

A schedule with colorful squares

Description automatically generated

Figure 3 Team meeting schedule (Bi Weekly all hands has been removed)

More specific meetings are scheduled on an ad-hoc basis through huddle on slack between relevant team members. Every few weeks the team will organize a design workshop where we will meet in person on campus to design and prioritize features for the product based on the information gathered from research, competitor analysis and user feedback.

Big decisions are made by consensus within the team unless there are instances where expert decision making is required. The role of scrum master/project manager is rotated weekly.

# Challenges:

## Integration of parts into one cohesive product

To mitigate issues of deployment of the final product, as reported by previous years teams, the team’s early objective was to deploy a minimal viable product (MVP) by week 4. While we were able to show each of the product’s components by this time and have some integration, we did fail in delivering a full stable, which has delayed our planned user feedback sessions.

This project has been a steep learning curve in the overall tech stack, and we have had to pivot early to try and achieve the MVP milestone. For example, we were unable to include the column specific cleaning functions created or the D3 dashboard in this release due to the integration issues. We switched from a python flask API backend to a microservice. Members of the data science team that were building the python functionality had to watch tutorials on YouTube on how to build a flask API as before this project the DS team were familiar with only python scripts in Jupyter notebooks.

@Naveen More on the challenges from your side.

## How to differentiate our product from competitors given limited time

Data cleaning as a product has been around for a long time with plenty of established competitors that offer a wide variety of functionality. During our research we have identified at least ten of them. For each of the competitor analysis we looked at key features, the pros and cons of the product and how we would differentiate. One of the key takeaways that we got from this was that we should build best in class visualizations. Those

## How to incorporate more complex data science techniques

Data cleaning itself is data science but while it is probably the most time consuming, most of it is also very basic and not too complex. The team have been working on finding the balance between keeping the product simple for the user but to also include some more advanced data science techniques to make this a worthy project of a data science masters. Machine or deep learning algorithms don’t really fit into the scope of data cleaning but advanced data science techniques such as PCA for dimension reduction do and we need to try and incorporate these techniques into the product.

# Timeline

After 6 weeks of working together we have a better idea on estimating time for tasks.

A screenshot of a computer

Description automatically generated

Figure 4 screenshot of Jira board

The below Gnatt chart shows the timeline for the entire project organised by tasks required for MVP, Beta, Alpha and general release. For the next two weeks to team will focus on the two big features needed, basic column cleaning and an interactive data profiling dashboard. Once these have been completed the team will then focus on building the more advanced data science functionality such as dimension reduction, creating new columns and statistical profiling of the user data. New features may be included or removed as the project continues pending user feedback from the MVP and Beta release. The last weeks will be focussed on a stable general release and collecting user feedback. A code freeze on new functionality will be in place from December first, focussing instead on stability and testing.

A screenshot of a computer

Description automatically generated

Figure 5 Gnatt chart for project completion

References

1. Srivastava, S. Bhardwaj and S. Saraswat, "SCRUM model for agile methodology," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 2017, pp. 864-869, doi: 10.1109/CCAA.2017.8229928.
2. Gaborova, M., Karuović, D., Kavalić, M., Radosav, D., Milosavljevć, D., Stanisaljev, S., & Bushati, J. (2021). Comparative analysis of agile and traditional methodologies in IT project management. *Journal of Applied Technical and Educational Sciences (jATES)*, 11(4), 1-24. <https://doi.org/10.24368/jates.v11i4.279>
3. Hu, Z., Yuan, Q., & Zhang, X. (2009). Research on Agile Project Management with Scrum method. 2009 IITA International Conference on Services Science, Management and Engineering.