TM351 Jupyter notebooks project Analysis of interviews

Sharon Dawes, Chris Thomson, Stephen Rice and Stephen Bowles

May 2022

Aims and scope of the project

The project investigated how Jupyter notebooks supported students' learning on TM351 Data Management and Analysis. It examined how students divided their time between theoretical reading on the TM351 module website and practical work using notebooks, both as a part of their learning, and then practice for their final report. It explored to what extent their use of the notebooks reinforced their learning or extended their knowledge and whether there was any relationship between where they were currently studying and what computing devices they had available when the student worked at home, at work or on the go that might have affected their ability to undertake practical work using notebooks as opposed to theoretical reading using other provided materials.

The research questions the project set out to answer were:

- Does the use of notebooks effectively support students in their studies of the module learning outcomes?
- Do notebooks allow the students to integrate theory and practice?
- Do students have technological problems in using notebooks?
- Does the use of notebooks dictate when and where students are able to study?
- Is the use of notebooks restricted by accessibility constraints?

This expanded on our previous survey, Dawes, Thomson, Rice and Bowles (2020, 2021) and Thomson and Dawes (2022).

Method

We conducted interviews with 10 students by using Skype for Business and analysed the transcriptions of these.

We asked students:

- how they studied in a typical session and how they integrated their learning on the practical parts of the module with the theoretical parts on the OU (Open University) website
- how they found information from previously studied topics on the module and how the notebooks enabled them to learn new techniques.
- how they used the notebooks in preparing assignment questions and to what extent the notebooks supported their report writing.

We also explored whether the students' places of study or computer equipment had any bearing on what they were telling us and whether the module would benefit from any additional resources.

Finally, we asked them whether they found the practical or theoretical work easier to study and whether there were any aspects of the module that led to confidence issues.

Analysis

Previous experience of students

The ten students we interviewed had very diverse previous experience. Some were new to the OU or were nearing the end of their first degree with us while others were already experienced software engineers or data scientists who had used Jupyter notebooks and virtualisation before. Others brought experience of a variety of external courses to their studies. All but two students had a good grasp of Python from M269 or other previous study, but a common theme was that this had not prepared them sufficiently for its use in combination with libraries on this module. Of the students without the prerequisite modules, one was working at a high enough level for this not to be a problem, having benefited from the basic Python familiarisation notebooks at the start, but the other appeared reported that the practical work was difficult.

Interleaving of resources

Seven of the students were starting with the practical work for a topic and following this up with the work on accompanying notebooks, sometimes returning to the theoretical work to consolidate aspects they had not understood.

Two students were doing practical work as soon as they reached relevant sections in theory, and one was basing his work on the notebooks. They found having multiple screens useful for this.

Students did not necessarily complete all the notebooks and one mentioned skipping those that were not needed to understand work for an assignment.

Problems with the technologies used on the module

While most students had no problems in installing the software, for a significant few there were difficulties in installing it from the supplied USB stick. For some this was overcome by a provided download. Configuration sometimes proved challenging in that the amount of RAM used by the software needed to be considered as a trade-off between resources allocated to the host system and the Virtual Box container. One student also had insufficient space on his hard disk and resorted to an external disk to solve this.

Students reported that notebooks were, in some cases, very long, especially those used for the EMA and if a notebook crashed it took a while to get it up and running again. The separate tabs opened up in students' browsers for each opened notebook could also overload their systems' resources.

All but one these problems were resolved using the technical help forum and students reported that the module-based instructions on how to get up and running were pretty good, though they would have liked a little more background at the start about why they were doing what they were doing to

create the environment. However, one student needed a lot of help from his tutor to resolve problems of the database notebooks not working.

A comment from one student was that he had tried to compile the software on his own server, and he would have liked this to have worked so he could have accessed this over his own network.

Where students studied

All students did some of their work at home. There was a rich variety of other places to study, ranging from hotel rooms on work trips to libraries, the pub and on public transport while commuting. However, working in the office was not popular. One student commented that doing university work in lunch breaks had never worked out as he would “never really retain the information because my mind wasn't in that place”. Another however said he would visit a website and read a few pages on his lunch break if he was behind.

What technology the students used

Of the students interviewed, three were using Macs, one Linux and the remainder were using Windows. Safari, Chrome, and Firefox were the browsers these students used. One student reported using three screens for his work so he could see practical activities, theoretical content, and his own notes at the same time.

While notebooks were of necessity studied online by all the students, four of the students were using mobile phones or tablets to read theoretical content when out and about and one had printed a huge amount to refer to even when at home.

One student referred to using a Git server to store all his work, so he had version control.

How the students integrated their knowledge of theory with practical learning

Students found some of the comments in the notebooks that included snippets of theory useful, but most would not have wanted the notebooks to have been any longer to include more of this. However, one student commented that the notebooks were distinct from the module material rather than intertwined in it and that there could have been more text in the notebooks to take you through.

However, one theme was that they really liked the way that the notebooks allowed you to describe some material, show a code snippet, run it, see what happened, and build up knowledge in that way.

For most students, their learning was driven by the assignments they needed to complete so that the assignment questions were relevant to the material they were studying.

One student mentioned building on the explorations in the notebooks to answer deeper questions and delving more deeply into provided bibliographies to expand his knowledge.

Searching for information on previous topics they had studied

A common theme was that it was difficult to search notebooks though they found searching the theoretical material on the module website easier. One student had printed all the materials to pdf and used a search on his laptop to find information on that using grep on his Linux PC. None of the Mac students mentioned using grep.

Students often resorted to using Google rather than looking for techniques in the previous notebooks unless they remembered which one to look in, though the notebooks were fairly well labelled to help with this, though there was a comment that Google had to be used with care to avoid using material that was not credible.

They said that in general, what was in the course materials was sufficient except for interest's sake, though a WhatsApp group was mentioned by one student that he used to talk through ideas with.

What was useful however was a resource written by a student consisting of summaries of each of the different formulae and codes used by the various notebooks. One student also referred to YouTube videos they had been alerted to on their Discord group that they had found very helpful for improving understanding of techniques.

How students took notes

Only one student was using Jupyter notebook for notes. Five students took no notes at all though one used Evernote to store results from running notebooks and sometimes inserted comments in these. The rest used a mix of handwritten and printed notes.

Of those who took notes, only two mentioned these as being very comprehensive. One who did not take notes used post-it notes on printouts, as well as Mind Maps, and a journal to record what he had tried, so this may have been equivalent.

Learning New Techniques

Students learned from executing code in the notebooks, sometimes splitting the code up so they could understand what each part was doing or changing the parameters to see the effects and for some, repeating the work later to consolidate learning.

Whilst the notebooks were seen as a really good learning tool that helped to cement knowledge, and the examples were seen as very useful, individual comments included that sometimes their scope was quite narrow, e.g., using only Milton Keynes for the maps.

An early notebook was seen as vital to understanding by a student who had experience only of other languages than Python and some notebooks were skipped altogether by a student with a lot of prior experience.

Which was easier, theory practical work?

Eight out of ten students thought that the practical work was easier. One student said this might be because he had already read the theoretical content. Later notebooks were perceived to be more difficult than earlier ones. The point was made though that it was useful to be able to pick up the theoretical work from anywhere whereas you needed to be at your computer to do the practical work.

How did they progress from provided notebooks to assignment ones

Students started by making copies of provided notebooks then copied and pasted code appropriate to the questions they were being asked either from other notebooks used Google to find this. Only one student talked about an iterative approach whereby they started with a subset of the data to check the output they were getting was meaningful.

Asking for help

Six students attended several or many tutorials and found these helpful. Two were not interested in these and the remaining two attended very few.

Only four students regarded their tutor as the first person they would go to for help and four students found the module forums useful as well as helping other students there.

Other sources of help mentioned were YouTube videos and Google.

How the notebooks supported report writing

Students found the notebooks very beneficial but there were mixed views on whether they were structured in the same way that a final report would be. A comment was that if you wrote the notebook in a very structured format in the first place so it flowed in some kind of chronological start to finish order it could be a very useful and powerful tool with the ability to combine both text and data analysis in one place.

Submitting the Word or pdf document with the notebook as supporting evidence was seen as the correct thing to do. One student however wrote this in LaTeX.

A disadvantage of the notebook for the EMA was that it was very long.

What affected their confidence in studying the module

Most students reported few confidence issues. Those that were mentioned included:

- The workload on the module
- The number of tabs that ended up being opened at once for the notebooks
- The context of some of the data being out of their comfort zone
- The phrasing of one TMA (tutor marked assignments) question where data was incomplete for one nation
- The steepness of the learning curve at the start

What future students would find useful

Amongst comments for improvements were that they would have liked:

- A bit more background on how the notebooks work
- Some Python consolidation materials before module start
- More coverage of basic Pandas and SQL
- A resource summarising keyboard shortcuts
- More examples for e.g., graphs and maps
- More assignments would have been welcome for one student or for another, smaller tests and quizzes in addition that would help embed the learning
- A way of indexing the cells as the numbering was not persistent
- Some guidance on how to make the notebooks look good (for reports) and how to create a notebook that pulled in other notebooks easily and ran all joined up and connected

Other comments about notebooks

Overall, the students really liked the module

A typical comment was that they liked "the fact that the notebooks were immediate, you didn't have to sort of mess around getting set up with like an IDE or compiling code or anything like that, it more or less just ran, you know, you can type a line of code and then you get an immediate response. And the same with the visualisations and stuff like that. So, I found that really useful."

Conclusion

Overall, the students enjoyed the module, and they especially liked the fact that the notebooks provided immediate results without compilation of code and the visualisations they generated.

References

Dawes S.R., Thomson C.D. (2019) "An investigation into the way Jupyter Notebooks enhance learning and teaching on TM351". Available at: http://www.open.ac.uk/about/teaching to and to learning/esteem/sites/www.open.ac.uk.about.teaching to and to learning.esteem/files/files/2019 to 10 to Dawes to Thomson to poster.pdf

Dawes S.R., Thomson C.D., Rice, S. & Bowles, S (2020) "How Jupyter Notebooks enhance learning and teaching on TM351", presentation at *eSTEeM conference 2020*. Available at: http://www.open.ac.uk/about/teaching to and to learning/esteem/sites/www.open.ac.uk.about.teaching to and to learning.esteem/files/files/2020 to 04 to 29 to 30 to Parallel to Session to D to Sharon to Dawes to et to al.pptx

Dawes S.R., Thomson C.D., Rice, S. & Bowles, S (2021) "An investigation of how Jupyter Notebooks enhance students' learning of data management and analysis", *Horizons in STEM 2021 Conference Proceedings*, pp.30-31. Available at: https://www.open.ac.uk/about/teaching-and-learning/esteem/sites/www.open.ac.uk.about.teaching-and-learning.esteem/files/files/2021-06-Horizons-in-STEM-HE-Conference-Proceedings-FINAL.pdf

Kluyver, T. et al., (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt, eds. Positioning and Power in Academic Publishing: Players, Agents and Agendas. pp. 87–90.