MEI Introduction to Data Science pilot

MEI has developed a bank of resources to introduce A level Mathematics students to Data Science. This is aimed at students who have studied the first year of A level Mathematics and can be completed in the summer term of year 12 or over the summer break between years 12 and 13. Over 100 schools and colleges requested access to the pilot materials in summer 2020. For more details see: *mei.org.uk/data-science*

Aim of the materials and main concepts

The aim of the materials is to give students a basic understanding of data science and a practical introduction to its techniques and applications so that they can make informed choices about further study and careers.

Main concepts:

- Data Science involves gaining knowledge from data sets through analysis and interpretation. These data sets are often very big and can involve live data.
- The pre-processing and cleaning of data (data wrangling) is a major part of a data scientist's role.
- Data presentation and visualisation are central to gaining insights and knowledge from data. Data presentation is not limited to the standard palette covered in statistics.
- The use of coding is essential to automate many of the processes so that they are scalable to very big and/or live data sets and to underpin machine learning.

Design of materials

The materials are designed to be a short self-study course. The majority of schools in the pilot offered it as an optional course alongside A level studies in the summer term or the summer holidays. The course is structured into 8 lessons, each of which is intended to be approximately 1 hour of study plus some suggestions for further investigation/reading.

The materials contain the following resources:

- Short videos introducing the concepts.
- Coding activities (hosted on *www.kaggle.com*).
- Videos of practising data scientists discussing their work.

The materials use the A level *Large Data Sets*. These are compulsory data sets for A level Mathematics. The materials feature examples and activities based on all of the different data sets for the four main A level Mathematics specifications.

The materials are split into 8 lessons:

1. Introduction to Data Science
2. Pre-processing and cleaning data
3. Cleaning, formatting and grouping data
4. Data presentation/visualisation
5. Exploring association
6. The Data Science cycle
7. Introduction to machine learning
8. What's next with Data Science

Impact

All schools were surveyed in September 2020. 40 teachers from 39 schools completed the survey. A further 24 teachers contacted us to say they had not used the materials, mainly because of the impact of school closures due to Covid-19.

Based on the schools who responded to the survey, 414 students completed at least one of the lessons and at least 177 students completed all the lessons. The actual number is likely to be slightly higher than this as some students from schools that did not respond to the survey also engaged with the materials. 155 of the 414 students were female, this is similar to the proportion of students taking A level Mathematics that are female.

Other survey findings

The responses from teachers about the content and structure of the resources were very positive. Many teachers fed back about the general difficulties of setting work remotely and supporting students during lockdown. This also impacted on the ability of some teachers to accurately assess how much their students had engaged with the materials. However, there was a general appetite for providing an opportunity like this for students in the context of a more 'normal' school year. See Appendix 1 for full results from the survey for teachers.

Quotes from teachers and students:

> *"The materials were very well put together."* (Teacher)
>
> *"I'm really pleased this was created and next year we intend to use it again - hopefully in person/in school!" (*Teacher)
>
> *"I loved it!"* (Teacher)
>
> *"Very helpful and enjoyable for a range of abilities even for someone just trying to become more familiar python."* (Student)

Potential for a follow-on course

We intend to explore the potential for a longer course that would follow on from this. This could be a course that students take in the Autumn term of year 13.

In the survey of teachers we asked "Do you think any of your students would be interested in studying a further follow-on course that would run alongside their A levels in year 13?". The responses to this were: Yes (11), No (4), Not sure (25).

Next steps

- The resources will be updated based on feedback from schools. They will be and made freely available for use in summer 2021 by all centres offering A level Mathematics.
- A version of the course will be created to support the study of Core Maths qualifications.
- We will explore the potential for a longer data science course that can be studied in the Autumn term of year 13.

Tom Button
tom.button@mei.org.uk
October 2020

MEI Introduction to Data Science pilot

Appendix 1: Data Science pilot materials survey results

130 schools were sent the materials in June and July. 40 teachers from 39 of these schools completed the post-pilot survey for teachers. 24 schools informed us that they didn't use the materials due to the general disruption experienced in summer 2020.

Usage

Questions 1-4 focussed on the usage of the materials.

Based on the 39 schools who responded to the survey:

- 414 students attempted at least one of the lessons.
- 155 were female. This is similar to the proportion of students taking A level Mathematics that are female.
- 177 students attempted all the lessons.

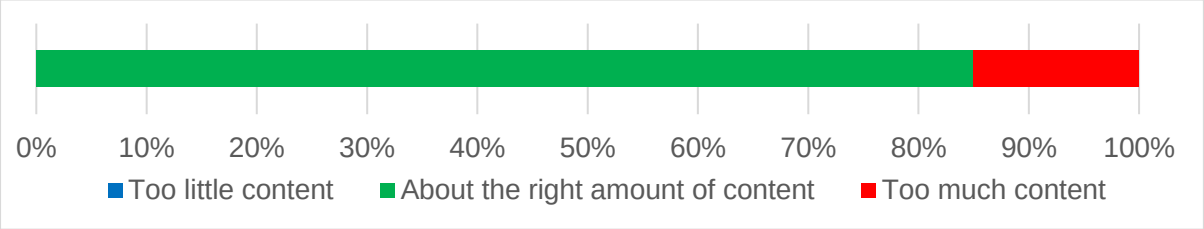
These values are based on the schools that responded to the survey. The actual number of students is likely to be slightly higher than this.

Additional comments about the types of students who engaged with the materials:

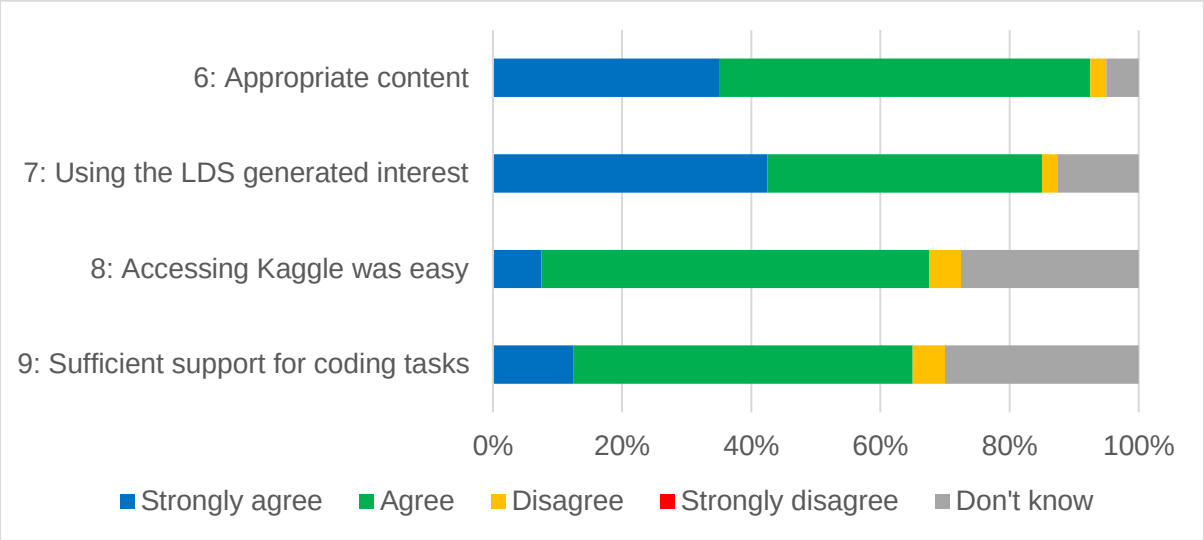
> *"A wider group of students initially expressed interest, but the ones who actually engaged with the materials were mainly those who were members of our Computer Science society and most were also studying Further Maths A-level."*
>
> *"4/12 of our maths students are female, all 4 completed the first 6 lessons."*
>
> *"Those with more of an interest in computer science. I had one student who is thinking of studying Data science at University."*
>
> *"High attaining students studying Further Maths. They didn't have any prior experience of coding but expressed an interest in Data Science as a potential career option."*

Content of the materials

5. The course was designed to be studied by year 12 students, between June and August, in addition to A level Mathematics. Was the content (8 lessons of 1 hour in length) an appropriate amount of study for the students:



6. The type of content (videos, coding activities, further reading) was appropriate to keep students' interest in the learning.
7. Basing the context on the A level Large Data Sets was useful in terms of getting the students interested in working on the materials.
8. The students found the coding activities on Kaggle easy to access.
9. The students were able to complete the coding tasks with the support given in the materials.

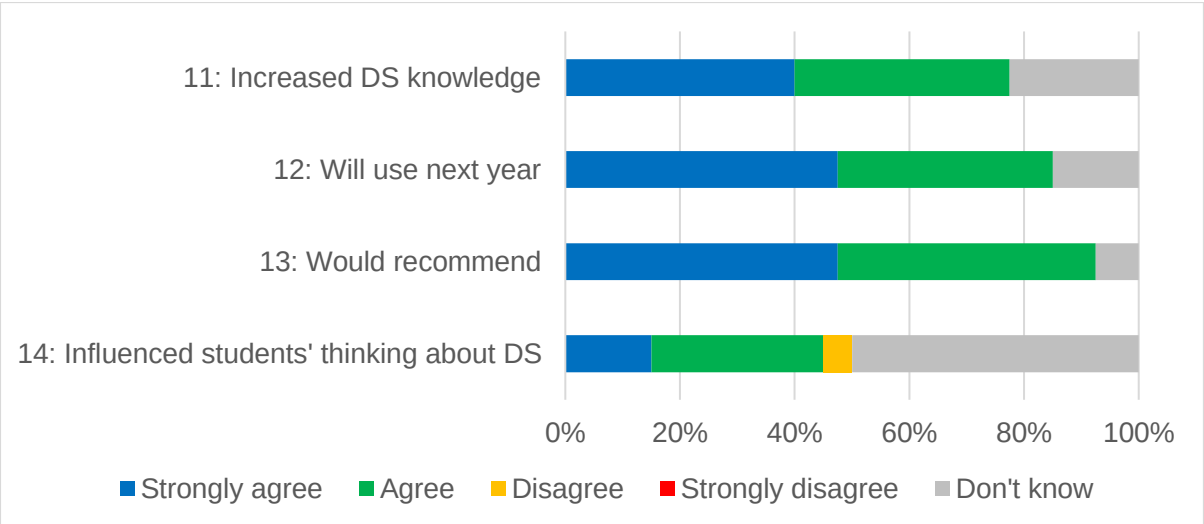


Selected general comments about the content of the materials (question 10):

> *"It may be better if students were able to access the material earlier, (maybe May) so that they will have more time to complete the tasks. In general the tasks were very good"*
>
> *"Coding bit was good if a bit slow to begin with"*
>
> *"It was a lot of work to expect them to study remotely on top of A level. I am interested in running a Data Science course in school over 6 weeks and then the content will be about right."*
>
> *"I loved it!"*

General comments about the materials

11. I believe that studying these materials has increased students' knowledge of data science.
12. I will use the updated version of these materials with my students next year.
13. I would recommend these materials to other teachers.
14. I believe that these materials have positively influenced my students' thinking about future study or careers in data science.



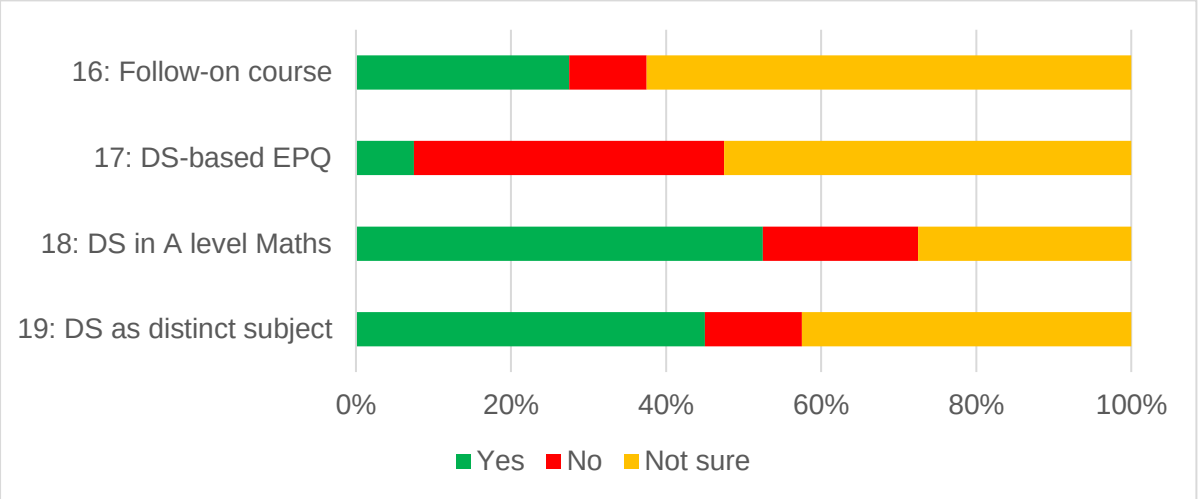
MEI Introduction to Data Science pilot

Selected general comments about the course (question 15):

> *"The materials were very well put together."*
>
> *"Really good linking careers to the large data set which is difficult to teach"*
>
> *"I'm really pleased this was created and next year we intend to use it again - hopefully in person/in school!"*

Potential for further resources/data science courses

16. Do you think any of your students would be interested in studying a further follow-on course that would run alongside their A levels in year 13?
17. Are any of your students considering using data science as a basis for an EPQ in year 13?
18. Do you think that if A level Mathematics should contain a data science element if it were to be revised at some point in the future?
19. Do you think that data science should be part of the school curriculum, as a distinct subject from Mathematics?



Other comments about the place of Data Science in the school curriculum (question 20):

> *"Perhaps offer two types of course. One more mathematical, aimed at A-Level Maths students. Another more general one permeating across the school curriculum."*
>
> *"Data science clearly plays such an important role in life and many careers these days, I would certainly like to see it firmly embedded in A-level studies (in maths or as a separate subject) and an option for high achieving GCSE students."*
>
> *"There's too much in A level maths now. Introducing data science would require use of laptops/desktops. We don't have them in school and our less well-off students generally have neither (they use phones/gaming machines)"*
>
> *"It could easily fit as a further maths module in the same was a Discrete or Decision mathematics. The old S3 used to have a coursework task in it and again Data could have a coursework/programming element to it."*