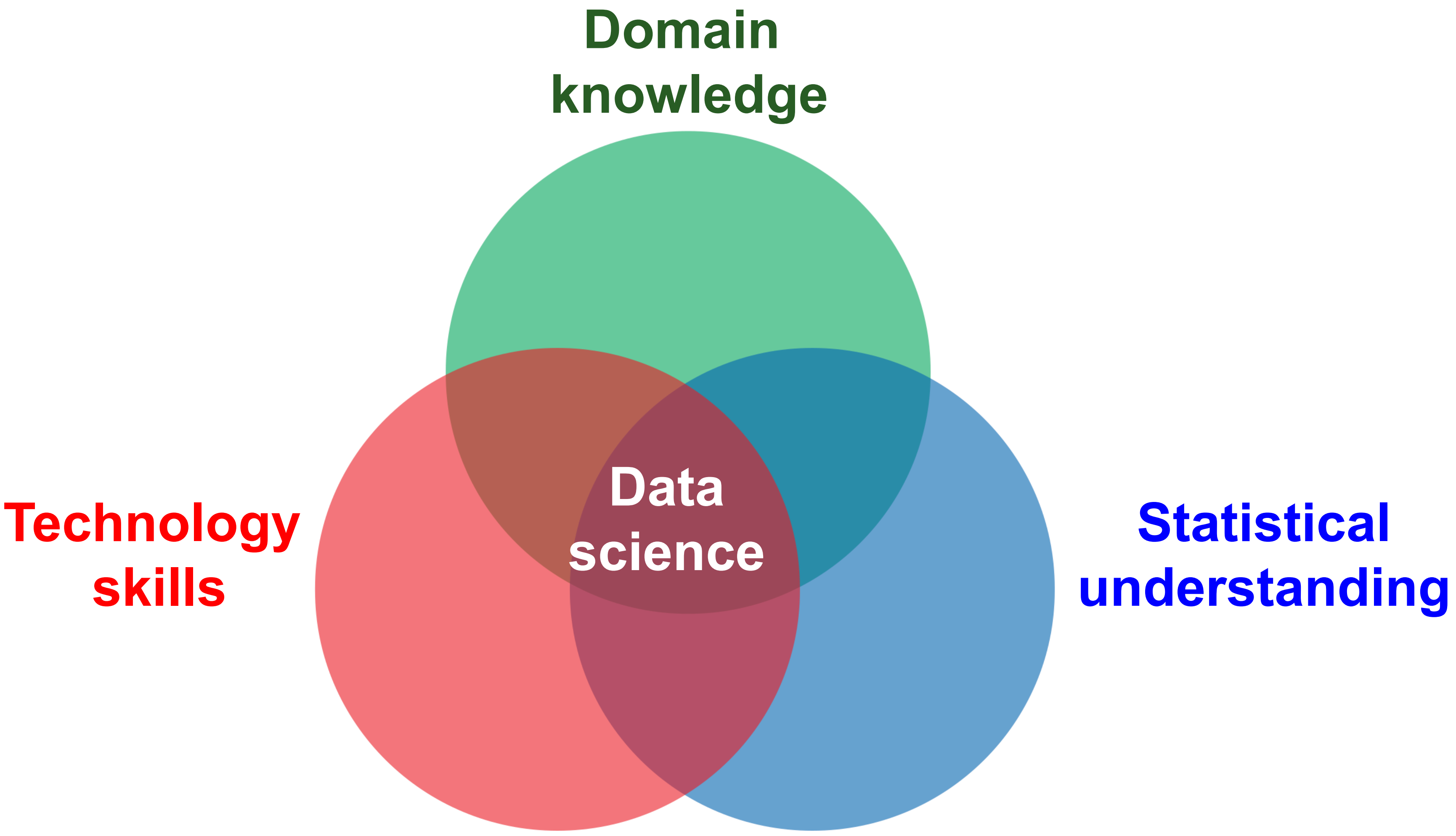
MEI Mathematics Education Innovation®

Design decisions in creating short data science courses for pre-university students

Tom Button
Ian Dickerson

Our definition of data science



Why is there a need for data science courses?

- Current school level maths and computing courses in England don’t contain sufficient data science skills

- Many students will progress to further study/careers where they will need to work with data

The space for Data Science

- **Long-term aim:**
  Data Science to be part of the school curriculum

- **Working towards this:**
  Optional courses for students to take alongside their *main* studies

MEI’s Data Science courses for students

Introduction to Data Science

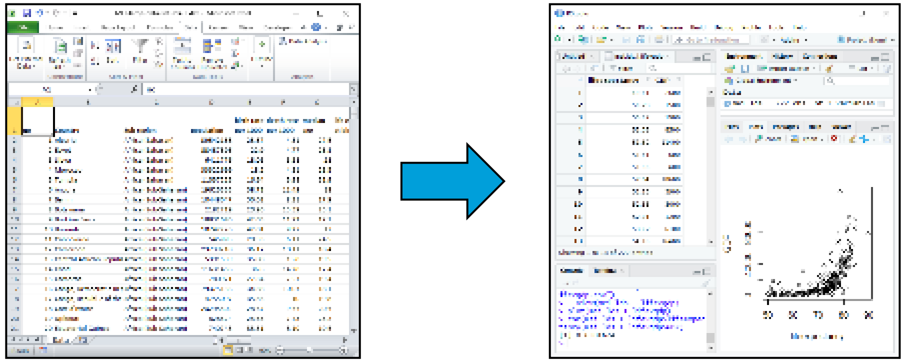
- Self-study
- 6 lessons
- Coding activities in Python
- Uses the *A level Large Data Sets*
- Not assessed
- Schools/colleges can award certificates

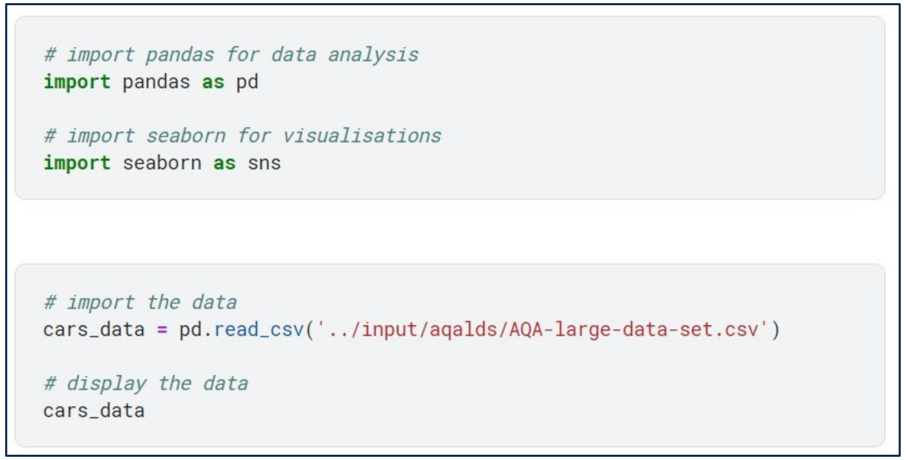
Data Science Taught Course

- Live online classes
- 10-12 lessons
- Coding activities in Python
- Uses a variety of contexts
- Assessed by a practical task and examination
- MEI certificate awarded

Aims of the courses

- Data sets need **pre-processing** and decisions are **context-dependent**
- A **programming language** is an efficient tool for working with data
- **Machine learning** is used to build models from data





```
# import pandas for data analysis
import pandas as pd

# import seaborn for visualisations
import seaborn as sns
```

```
# import the data
cars_data = pd.read_csv('../input/aqalds/AQA-large-data-set.csv')

# display the data
cars_data
```

Design questions

1. How to develop students’ abilities to make decisions based on the **context** of the data?

2. How to use a **programming language** without the course feeling like a coding course?

3. How to introduce **machine learning** to students working at this level?

Example: Making context-based decisions

CaseID	MCZ_1	MCZ_2	MCZ_8	RSEX	AGEXr	Martstat3r
10294	2	2	3	1	4	3
10296	4	3	5	2	6	1
10297	6	8	8	1	5	1
10298	8	7	8	1	3	1
10307	10	10	1	2	5	3
10307	9	10	10	1	2	1
10317	6	99	7	2	2	1
11656	10	7	10	2	6	1
11656	5	7	10	2	2	1
11674	98	98	98	2	2	2
11676	7	7	10	1	6	2
11680	9	9	9	1	5	1
11705	8	8	10	2	1	2

MCZ_2:
Overall, to what extent feel things you do in your life are worthwhile? (0-10)

Marstat3r:
Marital status

1: Married/cohabiting

2: Single

3: Widowed/divorced/separated

Developing pre-processing skills

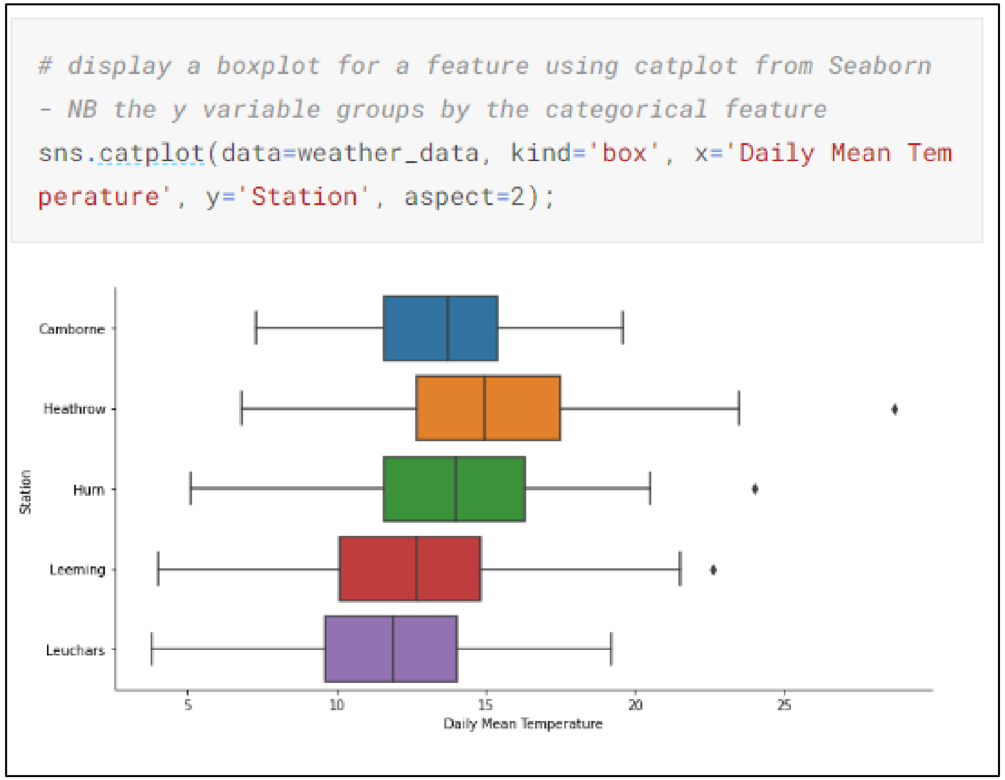
- The courses features a wide variety of different **real contexts**

- Students **see context-based decisions** in all the tasks but the tasks focus on analysis or modelling

- Some tasks include opportunities to **repeat pre-processing** techniques

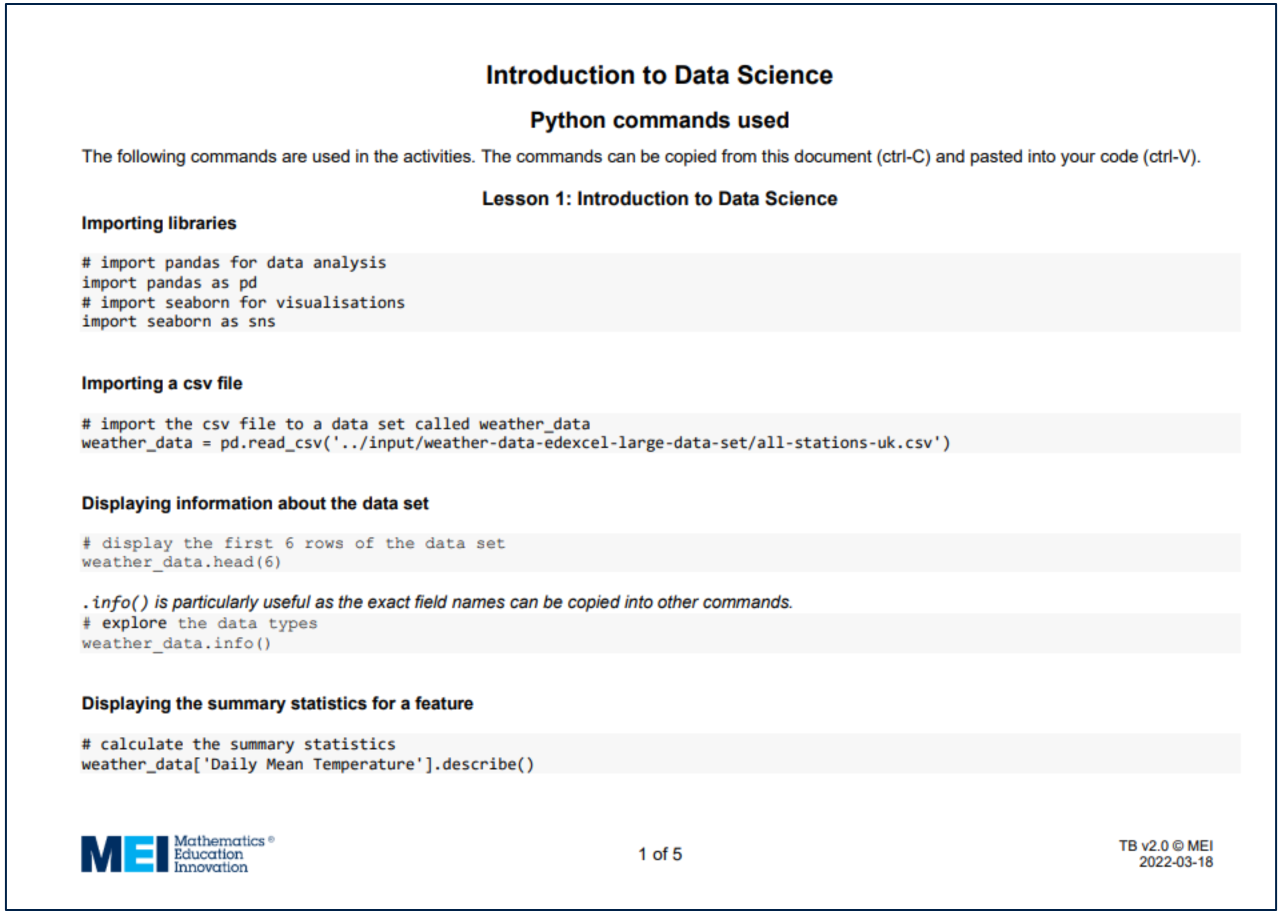
Embedding programming

- Coding skills are embedded in the tasks using Python notebooks
- All required coding commands are given
- Students are expected to copy and edit existing code



```
# display a boxplot for a feature using catplot from Seaborn
- NB the y variable groups by the categorical feature
sns.catplot(data=weather_data, kind='box', x='Daily Mean Tem
perature', y='Station', aspect=2);
```

All the commands used in the course are given



Introduction to Data Science

Python commands used

The following commands are used in the activities. The commands can be copied from this document (ctrl-C) and pasted into your code (ctrl-V).

Lesson 1: Introduction to Data Science

Importing libraries

```
# import pandas for data analysis
import pandas as pd
# import seaborn for visualisations
import seaborn as sns
```

Importing a csv file

```
# import the csv file to a data set called weather_data
weather_data = pd.read_csv('../input/weather-data-edexcel-large-data-set/all-stations-uk.csv')
```

Displaying information about the data set

```
# display the first 6 rows of the data set
weather_data.head(6)
```

`.info()` is particularly useful as the exact field names can be copied into other commands.

```
# explore the data types
weather_data.info()
```

Displaying the summary statistics for a feature

```
# calculate the summary statistics
weather_data['Daily Mean Temperature'].describe()
```

Mathematics® Education Innovation

1 of 5

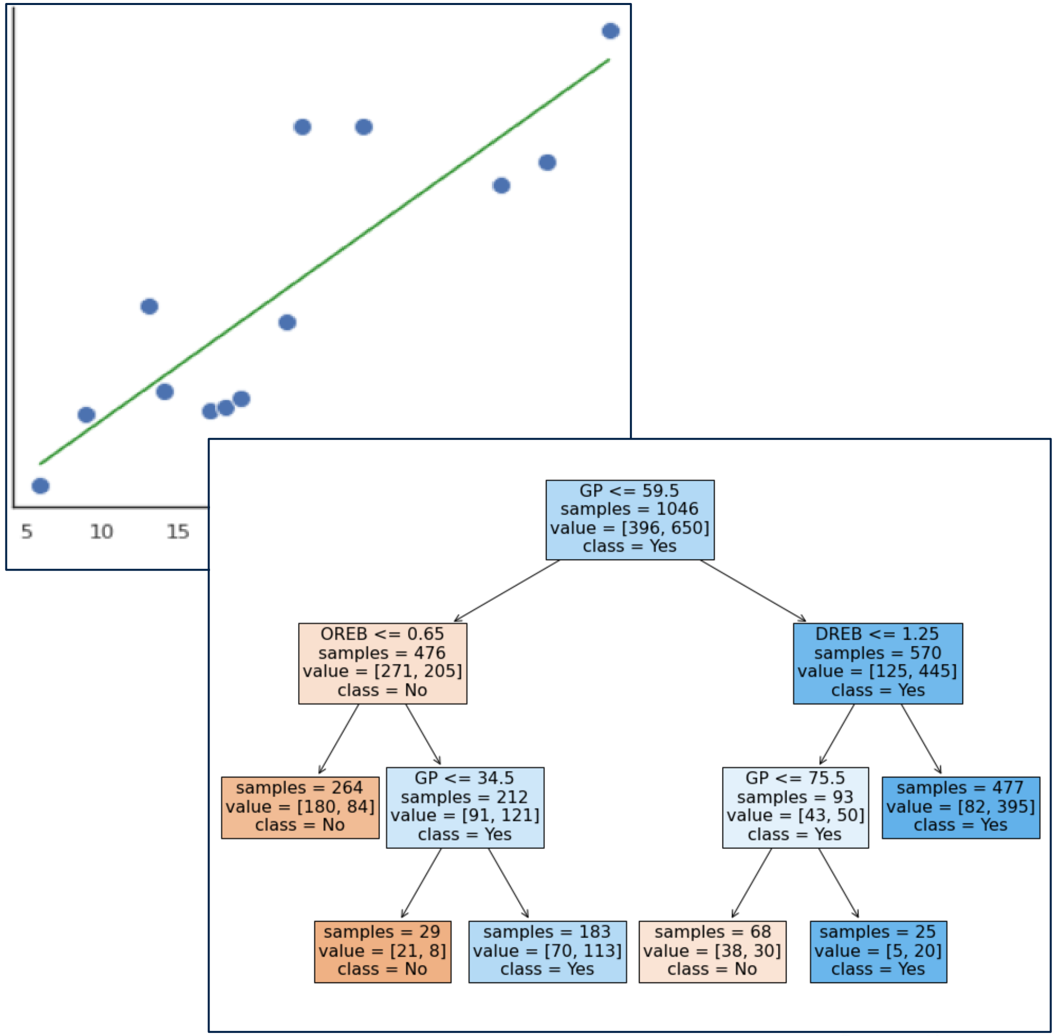
TB v2.0 © MEI
2022-03-18

Introducing machine learning

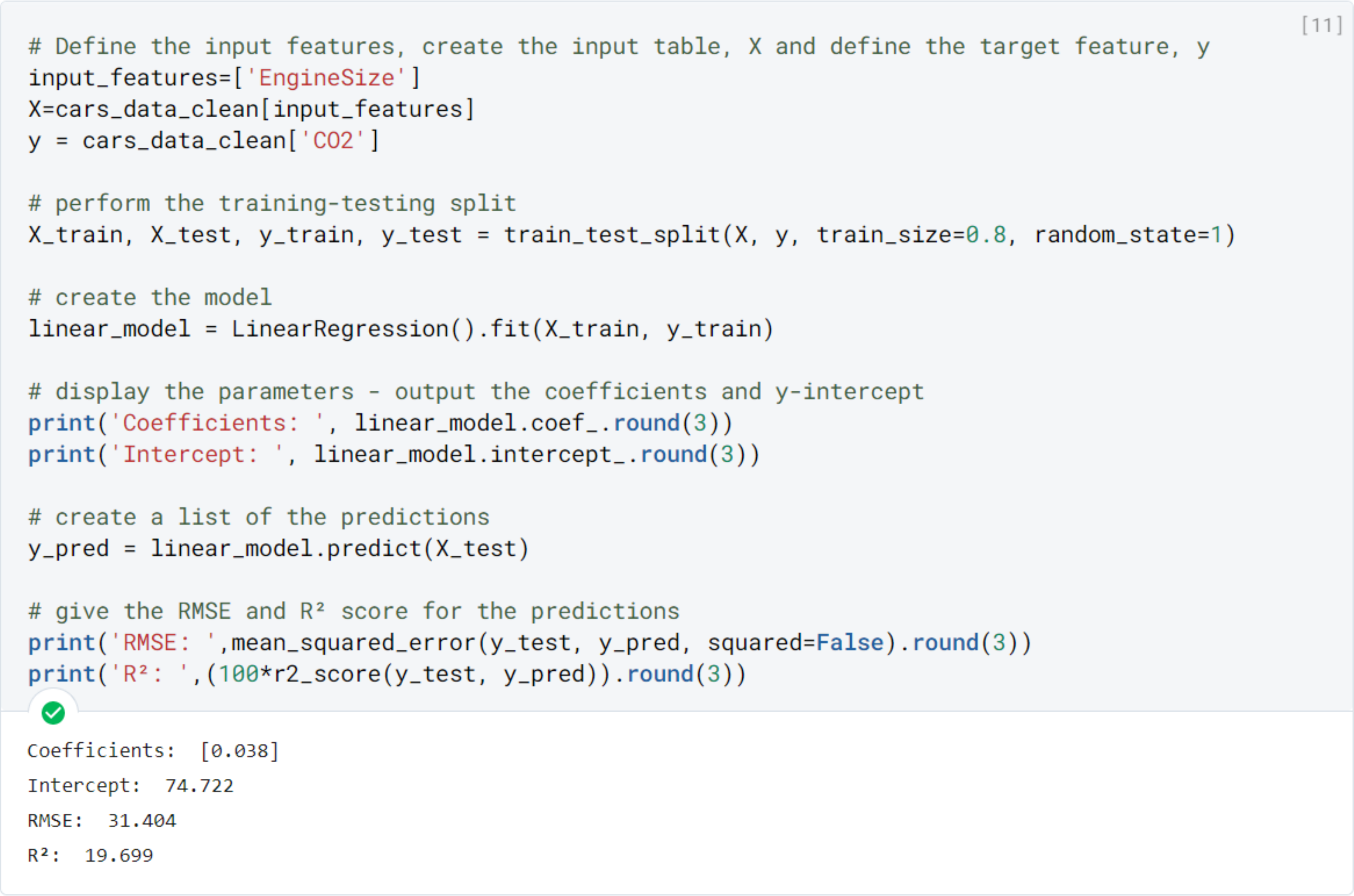
The focus is on the use of machine learning to build a **predictive model**.

Learners:

- Perform training-testing splits
- Use metrics to compare models
- Identify possible sources of bias in the data/model



The machine learning algorithms are treated as ‘black box’



```
[11]
# Define the input features, create the input table, X and define the target feature, y
input_features=['EngineSize']
X=cars_data_clean[input_features]
y = cars_data_clean['CO2']

# perform the training-testing split
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.8, random_state=1)

# create the model
linear_model = LinearRegression().fit(X_train, y_train)

# display the parameters - output the coefficients and y-intercept
print('Coefficients: ', linear_model.coef_.round(3))
print('Intercept: ', linear_model.intercept_.round(3))

# create a list of the predictions
y_pred = linear_model.predict(X_test)

# give the RMSE and R² score for the predictions
print('RMSE: ',mean_squared_error(y_test, y_pred, squared=False).round(3))
print('R²: ',(100*r2_score(y_test, y_pred)).round(3))
```

```
Coefficients:  [0.038]
Intercept:  74.722
RMSE:  31.404
R²:  19.699
```

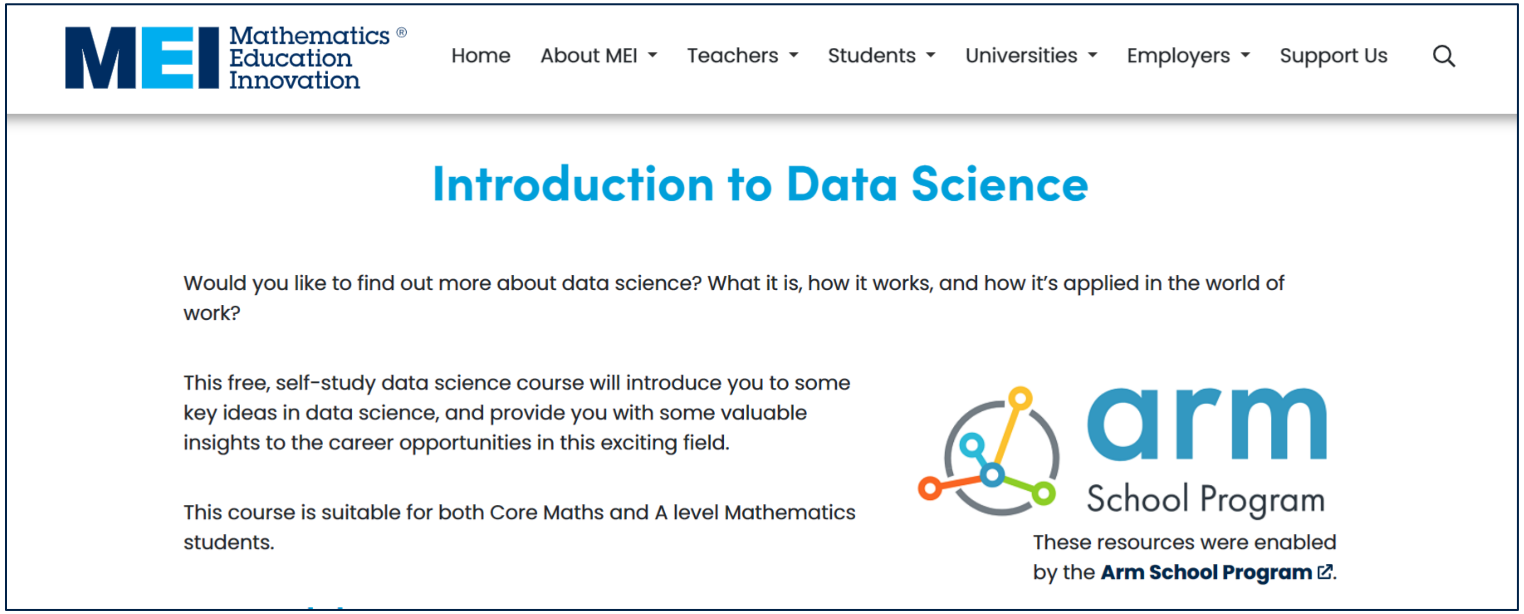
Design issues

1. How to develop students’ abilities to make decisions based on the **context** of the data?

2. How to use a **programming language** without the course feeling like a coding course?

3. How to introduce **machine learning** to students working at this level?

Accessing the materials



Introduction to Data Science

Would you like to find out more about data science? What it is, how it works, and how it's applied in the world of work?

This free, self-study data science course will introduce you to some key ideas in data science, and provide you with some valuable insights to the career opportunities in this exciting field.

This course is suitable for both Core Maths and A level Mathematics students.

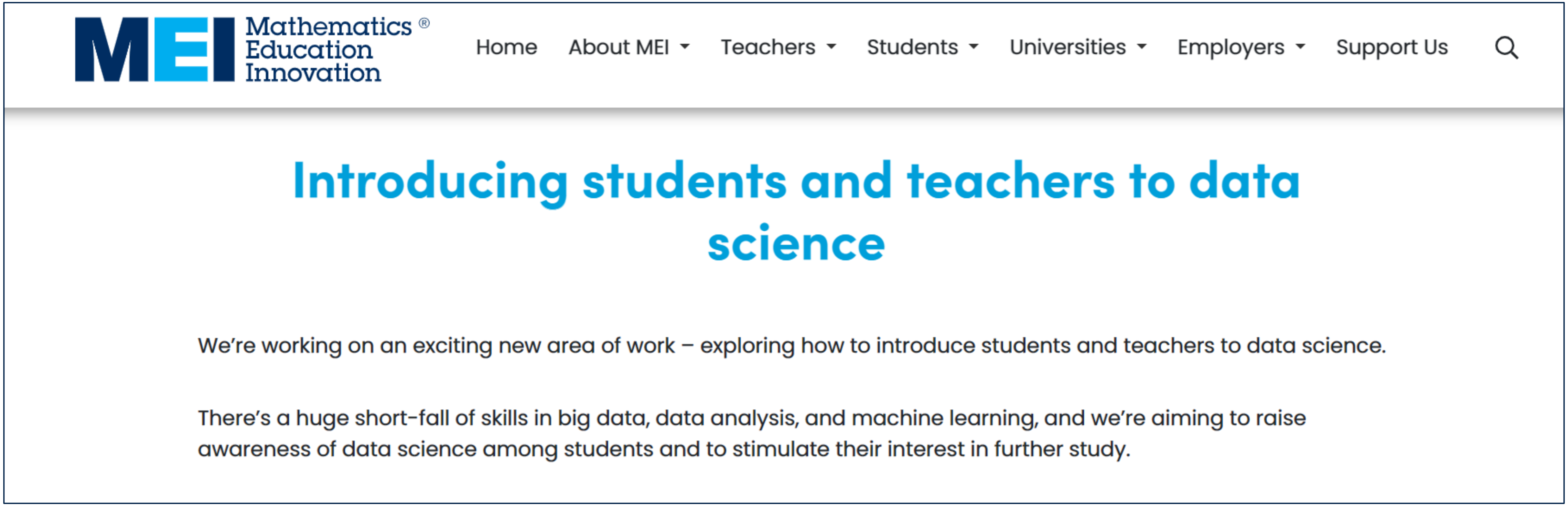
These resources were enabled by the **Arm School Program**.

mei.org.uk/introduction-to-data-science/

Students > A level Mathematics > More Maths

Students > Core Maths > More Maths

More information about MEI’s Data Science work



Introducing students and teachers to data science

We’re working on an exciting new area of work – exploring how to introduce students and teachers to data science.

There’s a huge short-fall of skills in big data, data analysis, and machine learning, and we’re aiming to raise awareness of data science among students and to stimulate their interest in further study.

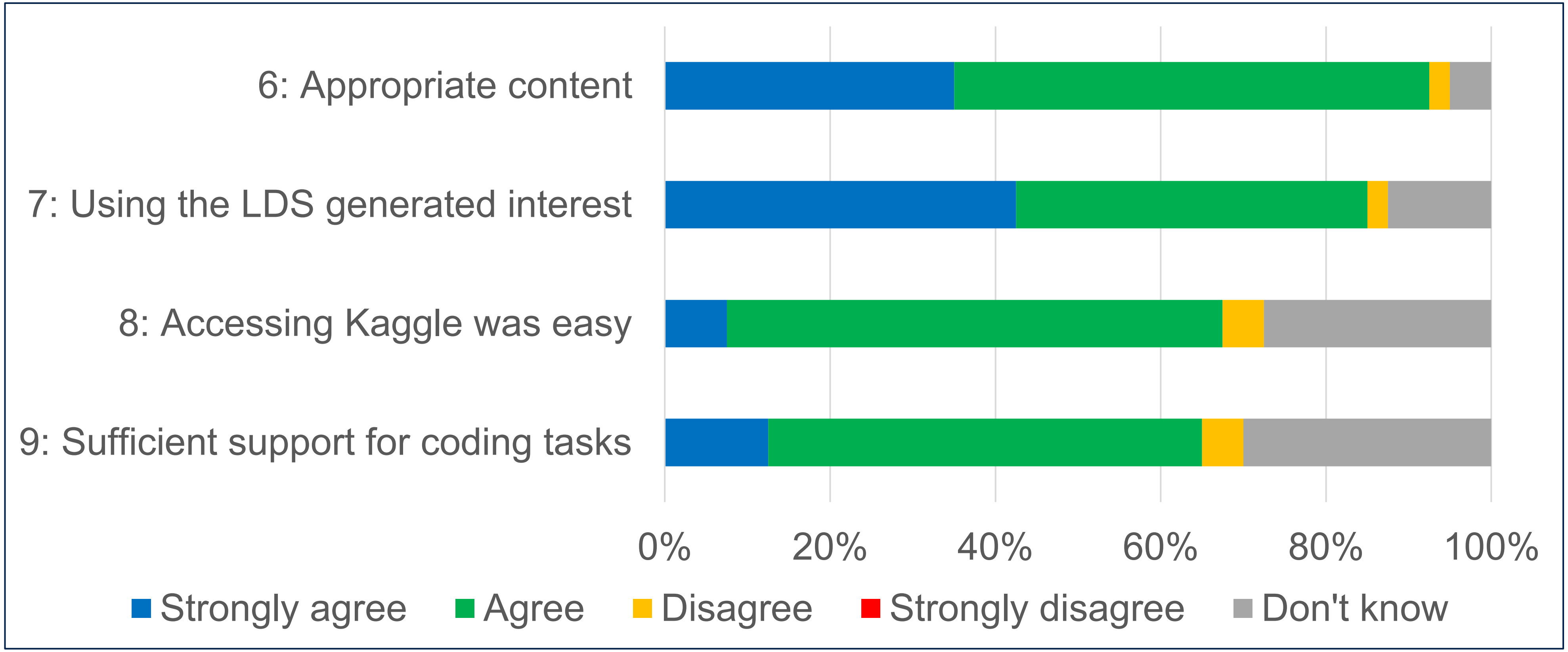
mei.org.uk/about-mei/what-we-do/current-projects-and-programmes/introducing-students-and-teachers-to-data-science/

Success?

- A few hundred students have used the short self-study course for each of the last three years

- In 2021-22 around 200 students signed-up for the taught course and 60 of these completed the assessments

Feedback on the pilot



Sponsored by
CASIO®

Contact details

Tom Button

tom.button@mei.org.uk

🐦 @MathsTechnology

Ian Dickerson

ian.dickerson@mei.org.uk

🐦 @mathsian