

# Applied Data Science

Introduction to Semester 2

Ken Cameron

# Applied Data Science

- Lecturers
- Structure
- Content
- Assessment

# Lecturers (recap)

- Ken Cameron
  - kmc55@bath.ac.uk
  - 1W 3.53 (Calendar on my door.)
- Adwait Sharma
  - as53339@bath.ac.uk
  - Wessex House 9.11
- Email is the best way to contact us directly.
- Moodle forums are better for general questions.

# Structure – Semester 1 (recap)

- Two weekly lectures
  - Tuesdays 1.15pm, 5W 2.3 and Wednesdays 9.15am, Edge Theatre
  - Weeks 1-5, 7-11
- Assessment
  - One lab exercise. (15%)
  - Exam (35%)
  - Tutors available for support in 2hr labs:

Check your timetable for which session you are allocated.

# Marking Update

- Marking of the S1 coursework is nearly done.
  - Sorry for the delay.
  - We expect to release the coursework marks this week.
- Exam marks will be released at the same time as S1 only units.
  - But, these will be provisional as they will not be reviewed by the Exam Board until the unit is finished in S2.

# CM50266 – Semester 1

Week	Week Starting	Lecture Tues 5W 2.3, 1.15pm	Lecture Weds Edge Theatre, 9.15am	Course Work Deadlines Always Friday 8pm
1	3 Oct	<b>Introduction (K)</b>	Data Representation + Ingestion (K)	Lab Release
2	10 Oct	Data Cleaning (K)	File Types (K)	
3	17 Oct	Intro to Visualization (A)	Dashboards and Infographics (A)	
4	24 Oct	Tools for Creating Data Visualizations (A)	Storytelling with Data Visualization (A)	Lab Data Collection Deadline
5	31 Oct	Big Data I (K)	Big Data II (K)	
6	7 Nov	<i>Consolidation Week – Labs Only, No Lectures</i>		
7	14 Nov	Data Protection (K)	Case Study – Data Review (K)	
8	21 Nov	Bias (K)	Case Study – Modelling (K)	Lab Main Deadline
9	28 Nov	Ethics (K)	Case Study – Research (K)	
10	5 Dec	Peer Assessment Guidance (K)	Case Study – Presentation (K)	
11	12 Dec	Review (K)	No Lecture	Lab Peer Forms Deadline

# Structure – Semester 2

- Mainly focused on the coursework. No new taught material lectures.
- Still some lectures, but:
  - Guest lectures from actual data scientists.
  - Less frequent and mostly in the second half.
  - Details announced when confirmed.
- Assessment
  - Case Study (50%).
  - Lab Sessions for support from Tutors.

# S2 Timetable

- Lecture slot

- 3E 2.4                      11.15pm                      Mondays.

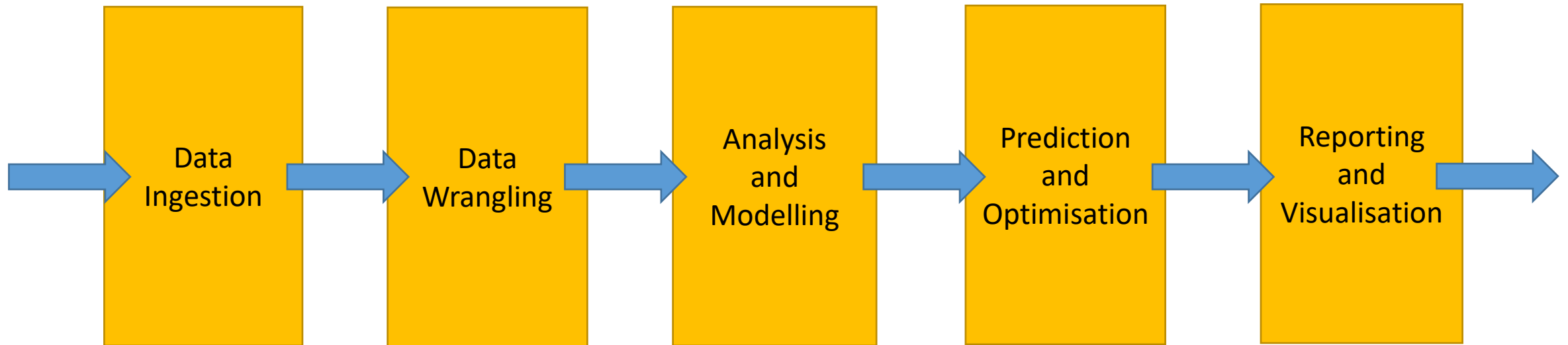
- Lab slots

- 10W 0.02                      9.15am - 11.15am                      Wednesdays.

- EB 0.8                      4.15pm - 6.05pm                      Wednesdays.



# Coursework: Data Science Pipeline Case Study



# Topic: Grocery Buying and Income

- You will use multiple datasets to answer the question:

“Do grocery shopping choices change with income?”

- You'll start with the Tesco Grocery 1.0 Dataset

# Tesco Grocery 1.0 Dataset

- Covers sales to Loyalty Card holders for the whole of 2015 across Greater London.
- Sufficiently granular to allow comparisons at a useful scale with other factors.
- Find details here:

<https://www.nature.com/articles/s41597-020-0397-7>

# Case Study

1. Explain what the dataset contains. What assumptions/limitations do we need to consider?
2. Provide a useful visualisation of the data to help identify any differences between areas.
3. Find income data at a suitable area scale over a number of years to allow comparison of the two datasets. Analyse this to identify any trends. e.g. areas on the up, in decline.
4. Use the two datasets to identify any patterns in buying behaviour. Do they correlate?
5. Present your conclusions as a video presentation and an executive summary.

# More skills than just CM50266

- As well as the material that Adwait and I have covered in S1, we expect that you will also make use of the content of other units.
- E.g. Statistics, Machine Learning.

# Initial Submission

- You will submit two items for assessment of the case study:
  1. A short briefing document (1 page.)
  2. And a 10 min (maximum) video presentation.

These are the only items that will form part of the assessment. (You will also submit any code you develop, but this will not be assessed, it is for reference only.)

# Peer Review

- As with the S1 coursework there will be a peer review exercise.
- You will only be reviewing the executive summary and video.
  - Any submitted code will not be shared and will not be part of the review.

# Use of Python Libraries

- In this assessment, you will likely need to do some coding.
- You are free to use existing libraries/tools unless explicitly told not to for a specific task.



# Use of AI tools



- We want to assess your work not ChatGPT.
  - Your use of AI tools should be limited to improving your work not creating it.
  - If you use AI tools, you should note this in the submission.
  - Remember the output of these tools are ‘hallucinations’.
    - You will need to check them.

# Q&A