CM52055 – Analytic software technologies

Coursework: Data Science Project

Set: 04/11/2025 (Week 6)

Due: 05/01/2026, 8pm

Percentage of overall unit mark: 100%

Passing mark: 50**%**

Submission Location: Moodle

Submission Components:

Submit exactly 3 files:

1. Jupyter Notebook (**which must have all the cells displaying results / outcomes**)
2. JSON file containing answers.
3. Written report in a single document (PDF file)

Submission Format:

The 3 files must be the following formats:

1. Jupyter Notebook (.ipynb format)
2. JSON Results file (.json format)
3. Report (PDF format)

Anonymous Marking: Yes (please do not include your name in the report or filename)

Generative AI Assessment Categorisation:

1. Type A [Guidance] (**no GenAI use is allowed**)

1 Overview/General Information/Learning Outcomes

Please read this coursework specification carefully in its entirety.

This coursework is worth 100% of total mark for this unit. The coursework consists of two components:

1. **Code development (40/100 marks)** – by completing the tasks listed in the Jupyter Notebook and submitting the completed Jupyter Notebook (**which must have all the cells displaying results / outcomes**) and the answers for all tasks (as a JSON file) as per the instructions in the Notebook.

2. **Written report (60/100 marks)** – as per the instructions detailed in this document.

This coursework aims to give you experience working on a real-world project. You are provided with a recent dataset and guiding questions. Informed by your analysis, you will make decisions and ultimately answer the business questions and present your recommendations. The coursework is intentionally designed to lead you through the often-unclear stages of data analysis: starting with broad questions, formulating your own deeper lines of inquiry, and then seeking patterns and insights to make evidence-based decisions and conclusions. **This spirit of independent inquiry is central to the coursework and will be rewarded.**

Scenario

You have been hired as a Data Scientist by a UK university that plans to launch a new degree programme OR revamp an existing degree programme. The university seeks evidence on commercial feasibility, student demand, and the competitive landscape. As part of this project, you will analyse the National Student Survey (NSS) 2025 dataset (survey conducted in April 2025). Working from broad questions set by the Executive Board, your task is to explore the dataset, generate insights, and write an executive summary (your report) that contains your reflections and recommendation for the new OR existing degree programme based on evidence.

This is your first role as a Data Scientist. Use this project to demonstrate rigorous, independent inquiry, justify your approach, look beyond the guide questions, and present an analysis backed by transparent, reproducible evidence.

Dataset

For this coursework, you are provided with a pre-processed subset of the original NSS 2025 dataset. The details of this dataset are outlined in the Jupyter notebook along with the schema for the dataset.

2 Tasks & Requirements for the Coding Component (40/100)

You are provided with a Jupyter notebook which contains the necessary code skeleton. The notebook outlines 16 tasks and marks for each task. Your task is to carry out the data analysis by providing the code for 16 tasks. The tasks are designed to help you familiarise yourself with the dataset and gain insights through in this process. As part of this coursework, you will discuss your implementation details, insights, and the justification for your decisions in your report. It is therefore advisable to look beyond the guide questions and explore the dataset so you can provide well-supported decisions and recommendations.

Please read the instructions in the notebook carefully — both while conducting your analyses and when submitting your Jupyter notebook and results for marking.

3 Tasks & Requirements for the Report (60/100)

A **summary** report structured in four sections. In each section, students will be assessed on clarity, depth of analysis, justification of choices, and the ability to integrate theory with practice.

You are to write the report as a data scientist who is summarising their findings and making recommendations for your employer's "C-Level" executives (or the Executives Board). The report should demonstrate critical thinking, technical awareness of how to generate insights from a dataset, and the ability to tailor recommendations to organisational needs. Your report must have the following sections:

Section 1: Description of Your Analyses (20 marks)

- Describe the data analysis process you carried out as part of learning and generating insights from the dataset. For example, you can include your justification for conducting your analyses, your reflections, and the analysis outcomes and their robustness or reliability.

Section 2: Recommending a Degree Programme & Improving the NSS Scores (10 marks)

- Outline your recommendation for launching (or revamping) the degree programme at your employer's university. It should include the proposed subject (*cah3_subject* – see the notebook). Outline your findings on identifying the areas for improvement in the NSS scores based on the NSS dataset for your employer.

Section 3: Commercial Feasibility & Strategic Opportunities (15 marks)

- Discuss the commercial feasibility of your recommended degree programme. For example, you can reflect upon the potential demand for the degree programmes based on relevant information in the dataset such as student populations, student satisfaction scores and identify the potential competitors (i.e., other universities) and relevant degree programmes they may already offer.

Section 4: Scaling up For Larger Dataset (15 marks)

- The original NSS 2025 dataset contains several millions of records. Describe your approach to scaling up your analysis if you were to conduct your analyses for the NSS datasets over the past two decades (note: the NSS was launched in 2005) which is likely to involve tens of millions of records. For example, you may wish to include a description of scalable algorithms, infrastructure, tech-stack and tools which could be useful when analysing large datasets.

4 Formatting Guidelines

- This is an anonymous coursework. Please do not include your name in the report.
- This coursework is a GenAI Assessment Category Type A (Guidance) and use of GenAI is **not allowed**. Please note that using GenAI for your coursework, may lead to Academic Misconduct investigation.
- Word Count: **1,500 to a maximum of 3,000 words** (excluding references). Font & Spacing: 12pt font, 1.5-line spacing, standard margins.
- File Format: Submit exactly 3 files as specified on the 1st page of this document.
- File Naming Convention: AST_CW_Data_Science_Report.pdf

Failing to adhere to the formatting guidelines (e.g., using font size 10, including your name) may lead to negative marking which may result in **deduction of a minimum of 1 mark and a maximum of 10 marks**.

5 Feedback

You will receive summative feedback on your work within 3 semester weeks of the submission deadline.

The following is a guideline for marks:

- Above 50% – Pass.
- 55-59% – Good.
- 60-69% – Excellent. (Merit range)
- 70%+ – Outstanding. (Distinction range)

6 Academic Integrity

For more details on academic integrity in the GenAI context, please see Academic Integrity and AI Guidance. Your work will be checked to ensure that you have not plagiarised. For more information about the plagiarism policy at the University see: https://library.bath.ac.uk/referencing/plagiarism.

Remember that published work that you refer to in your report should be clearly referenced in your text and listed in a bibliography section given at the end of your report. For more information see, https://library.bath.ac.uk/referencing/new-to-referencing.

7 General Information

Requests for extensions should be made to your PGT Director of Studies, Dr Fabio Nemetz (mapfn@bath.ac.uk)**.** Lecturers and tutors cannot approve extensions. Please make sure you are familiar with the department's coursework deadline extension policy.

It is your responsibility to check and correctly submit your work. Once you've submitted to Moodle you should **download your submission and check it**. Submission is anonymous, so do not include your names in your report, or file names. The fact that the submission is anonymous means we cannot contact you if there is a problem with a submission, so it is **very important to follow the instructions and check it**.