

**SUBJECT TO EXTERNAL EXAMINER APPROVAL**

<b>5DATA004W Data Science Project Lifecycle (2024/25)</b>	
<b>Module leader</b>	Dr Philip Worrall
<b>Units</b>	Individual Coursework
<b>Weighting:</b>	30%
<b>Qualifying mark</b>	30%
<b>Description</b>	Students will be assessed on their ability to develop and test a technical solution using appropriate tools and methodologies.
<b>Covered Learning Outcomes</b>	<p>LO4 Evaluate and demonstrate an awareness of competing development frameworks, software tools and methodologies; justify the approach taken based on a consideration of the alternatives, available skills mix and environmental constraints.</p> <p>LO5 Prepare a robust test plan; detailing the scope of each test, justification, approach taken, resources used and the schedule of intended test activities.</p>
<b>Handed Out:</b>	<b>25<sup>th</sup> March 2025</b>
<b>Due Dates</b>	1pm 6 <sup>th</sup> May 2025
<b>Expected deliverables</b>	<ol style="list-style-type: none"> <li>1. Video demonstration of the dashboard</li> <li>2. Link to the published Streamlit app</li> <li>3. Link to the GitHub repository with the final app source code</li> <li>4. Report including the test log</li> </ol>
<b>Method of Submission:</b>	Online via Blackboard
<b>Type of Feedback and Due Date:</b>	<p>Written feedback on the submission will be provided within 3 weeks after the submission (the mark and comments via BB Rubric).</p> <p><b>All marks remain provisional until formally agreed by an Assessment Board.</b></p>

## **Assessment regulations**

Refer to the course handbook or Part 3 of the university Academic Regulations handbook for a clarification of how you are assessed, penalties and late submissions, what constitutes plagiarism etc.

### **Penalty for Late Submission**

If you submit your coursework late but within 24 hours or one working day of the specified deadline, 10 marks will be deducted from the final mark, as a penalty for late submission, except for work which obtains a mark in the range 40 – 49%, in which case the mark will be capped at the pass mark (40%). If you submit your coursework more than 24 hours or more than one working day after the specified deadline you will be given a mark of zero for the work in question unless a claim of Mitigating Circumstances has been submitted and accepted as valid.

It is recognised that on occasion, illness or a personal crisis can mean that you fail to submit a piece of work on time. In such cases you must inform the Campus Office in writing on a mitigating circumstances form, giving the reason for your late or non-submission, and if applicable, the relevant documentary evidence with the form. This information will be reported to the relevant Assessment Board that will decide whether the mark of zero shall stand. For more detailed information regarding University Assessment Regulations, please refer to the following website:  
<http://www.westminster.ac.uk/study/current-students/resources/academic-regulations>

# INDIVIDUAL COURSEWORK

## OVERVIEW

In this coursework assignment, students will assume the role of a data scientist in-training working on behalf of the government.

Students will apply the skills and experience gained on the module to develop an interactive dashboard to analyse a government dataset. The purpose of the interactive dashboard is to provide insight for government officials and/or policy makers.

To ensure successful completion of the coursework, students will need to demonstrate an awareness of relevant development and testing methodologies.

This coursework **MUST** be completed **INDIVIDUALLY**.

## YOUR TASK

Select an open dataset from either [data.gov.uk](https://data.gov.uk) or [data.gov](https://data.gov).

Your selected dataset should contain a minimum of 100 observations (rows) and 3 or more columns (variables). Refer to the set of example datasets below as a guide for what is suitable.

Students are **NOT** permitted to work on the same dataset. You **MUST** confirm your choice with the module leader before undertaking the task. A **10-mark penalty** will be applied to students that don't seek the module leader's approval or where two or more students use the same dataset.

Create an interactive dashboard using the dataset you selected. Your dashboard should use data visualisation techniques to present key findings and enable users to interactively explore the data.

The dashboard **MUST** be implemented as a Streamlit app and make use of a GitHub repository to manage and track the development process. Your development **MUST** be appropriately tested and include a set of test cases and testing log.

You will need to produce a 5-minute video to demonstrate the dashboard. In the video you should demonstrate interactive elements, highlight key findings and insights.

You will need to submit a report documenting the development and testing process. You **SHOULD** use the provided template attached at the end of the coursework specification.

## Example Datasets

- 1) [Tour de France social and economic impact report results](#)
- 2) [Baby names for girls in England and Wales](#)
- 3) [Bristol Air Quality](#)
- 4) [School mode of travel](#)
- 5) [Electric Vehicle Population Data](#)
- 6) [US Food Imports](#)

## DELIVERABLES

The following deliverable should be submitted via the module's Blackboard site (submission links are available under section "Assessment and Submission Links") by **1pm 6<sup>th</sup> May 2025**

### Report including the test log

The report should document key phases of the development. This includes the dataset selected, aims and objectives, development methodology, the identification of functional and non-functional requirements, a set of 5 test-cases and a test log. The suggested word count for the report is 1500 words.

The front cover of your report **MUST** contain links to the following **THREE** items:

**1) Link to a video demonstration of the dashboard**

In the video you should demonstrate the interactive elements of the dashboard. You should highlight any important findings or insights from your data analysis. Your face and voice **MUST** be clear in the recording. The duration should not exceed **5 minutes**. Videos that exceed 5 minutes in duration will be awarded a mark based on the first 5 minutes.

**2) Link to the published Streamlit app**

You will need to publish the app from your GitHub repository.

**3) Link to the GitHub repository with the final app source code (summative)**

The repository **MUST** be made accessible to the public. You are not allowed to make any changes to the repository after you submitted your work.

Students **MUST** ensure that the links to the GitHub repository, Streamlit app and video demonstration work and are accessible to examiners until the 1<sup>st</sup> of August 2025.

**5DATA004W DATA SCIENCE PROJECT LIFECYCLE - INDIVIDUAL COURSEWORK - MARKING SCHEME****NOTE:** Marks will be allocated in relation to the correctness, completeness, and the quality of the answer provided.

Criteria	Mark Per Component	Mark Provided	Comments
<b>COMPONENTS</b>	<b>100 marks</b>		
<b>AIMS AND DEVELOPMENT METHODOLOGY</b>			
<ul style="list-style-type: none"><li>• Appropriate aims and objectives for the selected dataset <b>(5 marks)</b></li><li>• Justification for development methodology selected <b>(5 marks)</b></li></ul>	<b>10</b>		
<b>REQUIREMENTS</b>			
<ul style="list-style-type: none"><li>• Relevant and well-defined requirements are provided <b>(5 marks)</b></li><li>• An appropriate number of requirements <b>(5 marks)</b></li><li>• The set of requirements are classified <b>(5 marks)</b></li></ul>	<b>15</b>		
<b>STREAMLIT APP</b>  (0 marks will be awarded for this component if the link to the Streamlit app is missing or not accessible)			

<p><b>Quality of visualisation (10 marks)</b></p> <ul style="list-style-type: none"> <li>• Appropriate and relevant visualisations</li> <li>• Create/innovative presentation of data</li> <li>• Clear and easy to understand representation of data</li> </ul> <p><b>Interactivity (10 marks)</b></p> <ul style="list-style-type: none"> <li>• Appropriate interactive elements</li> <li>• Good user control over interactions</li> <li>• Clear responsiveness to user interactions</li> </ul> <p><b>Design and usability (10 marks)</b></p> <ul style="list-style-type: none"> <li>• Clear and easy to understand layout and intuitive navigation</li> <li>• Consistent design elements</li> </ul> <p><b>Functionality (5 marks)</b></p> <ul style="list-style-type: none"> <li>• Dashboard provides enough functionality to explore key insights</li> <li>• No errors</li> </ul>	35		
<p><b>VIDEO DEMONSTRATION</b></p> <p>(0 marks will be awarded for this component if the link to the video presentation is missing or not accessible)</p>			
<ul style="list-style-type: none"> <li>• Clear demonstration of the dashboard's functionality and interactivity (5 marks)</li> <li>• Key insights highlighted (5 marks)</li> <li>• Appropriate presentation technique (5 marks)</li> </ul>	15		
<b>VERSION CONTROL</b>			

(0 marks will be awarded for this component if the link to the GitHub repository is missing or not accessible)			
<ul style="list-style-type: none"> <li>• Appropriate number of commits <b>(5 marks)</b></li> <li>• Appropriate commit messages <b>(5 marks)</b></li> </ul>	<b>10</b>		
<b>TESTING</b>			
<ul style="list-style-type: none"> <li>• Test plan including a set of 5 test cases covering all functional requirements of the application <b>(10 marks)</b></li> <li>• Completed test log covering all test cases <b>(5 marks)</b></li> </ul>	<b>15</b>		
A <b>10-mark penalty</b> will be applied in the case that <b>TWO</b> or more students use the same dataset or do not confirm their choice of dataset with the module leader prior to undertaking the task.			
<b>END</b>			

## Report Template

Individual Coursework 5DATA004W Data Science Project Lifecycle

Name:

Student Id:

Dataset:

Link to Video presentation:

Link to Streamlit app:

Link to GitHub repository:

## Aims and Objectives

List the name and source of the dataset selected. It must be a dataset from either DATA.GOV or DATA.GOV.UK.

Write about the aims and objectives of this project. Also justify the key insights you selected.

## Development Methodology

Write about the choice of development method selected to conduct the work.

## Requirements

List the functional and non-functional requirements.

## Test cases

5 test cases

Document your test cases here. Use the template below:

#	TC1	Title:	Test case Title
Description			
Steps and input data		1. Lorem Ipsum...	
Dependencies			
Expected result			

## Test log

Add your test log here. Use the template provided.

TC	Date	Executed by	Actual result	Pass/Fail	Notes



