The Assignment

# Introduction

Completing the assignment should allow you to demonstrate to what degree you have been able to achieve the module’s Learning Outcomes

You will be required to create a report in the form of a Jupyter Notebook that will contain Python code.

**Please read every section of this document very carefully before starting on your work, apart from specifying the tasks and questions, it also contains information on how your submission will be marked, very important information about Academic Integrity and instructions on how to submit. Not following the guidelines and/or instructions will affect your grade.**

## Marking Scheme

The assessment will be marked against the university’s Common Marking Scheme (CMS)

<https://www.stir.ac.uk/about/professional-services/student-academic-and-corporate-services/academic-registry/academic-policy-and-practice/quality-handbook/assessment-policy-and-procedure/appendix-1-undergraduate-common-marking-scheme/>

The overall grade of your submission will consider six (6) criteria: Dataset complexity, Code Quality, Analysis Depth, Visualisation Quality, Story Telling, and Originality and Creativity.

The following is a summary of what you need to achieve to gain a grade in the major grade bands.

### 1st (High)

**Dataset Complexity:** Exceptionally rich dataset chosen with a clear rationale, showing deep engagement with the subject matter.

**Code Quality:** Code is exceptionally well-organised, efficient, and well-documented.

**Analysis Depth:** Analysis demonstrates a deep and nuanced understanding of the data, going far beyond surface-level observations.

**Visualisation Quality:** Visualisations are highly effective, well-designed, and convey complex insights clearly.

**Storytelling:** The data story is compelling, insightful, and expertly structured, demonstrating a high level of engagement with the data.

**Originality and Creativity:** Work demonstrates a high level of originality and creativity in approach and interpretation.

### 1st (Mid)

**Dataset Complexity:** Rich dataset chosen with a clear rationale and engagement with the subject matter.

**Code Quality:** Code is very well-organised, efficient, and documented.

**Analysis Depth:** Analysis demonstrates a strong understanding of the data and thoughtful interpretation.

**Visualisation Quality:** Visualisations are effective and well-designed, conveying insights clearly.

**Storytelling:** The data story is engaging, well-structured, and demonstrates thoughtful engagement with the data.

**Originality and Creativity:** Work demonstrates originality and creativity in approach and interpretation.

### 1st (Low)

**Dataset Complexity:** Adequate dataset chosen with a reasonable rationale.

**Code Quality:** Code is organised and documented.

**Analysis Depth:** Analysis demonstrates understanding of the data and some thoughtful interpretation.

**Visualisation Quality:** Visualisations are mostly effective and convey insights.

**Storytelling:** The data story is reasonably engaging and structured.

**Originality and Creativity:** Some evidence of originality and creativity in approach and interpretation.

### 2:1

**Dataset Complexity:** Dataset chosen is somewhat basic, rationale is somewhat clear.

**Code Quality:** Code is somewhat organised and documented.

**Analysis Depth:** Analysis demonstrates a basic understanding of the data.

**Visualisation Quality:** Visualisations are basic but convey some insights.

**Storytelling:** The data story is somewhat engaging and structured.

**Originality and Creativity:** Limited evidence of originality and creativity in approach and interpretation.

### 2:2

**Dataset Complexity:** Dataset chosen is somewhat basic; rationale is somewhat clear but lacks depth.

**Code Quality:** Code is somewhat organised but may lack sufficient documentation or efficiency.

**Analysis Depth:** Analysis demonstrates a basic understanding of the data, but insights are somewhat superficial.

**Visualisation Quality:** Visualisations are adequate but may lack clarity or effectiveness in conveying insights.

**Storytelling:** The data story is present but lacks engagement or a well-structured narrative.

**Originality and Creativity:** Limited evidence of originality and creativity in approach and interpretation, with reliance on common or straightforward approaches.

### 3rd

**Dataset Complexity:** Dataset chosen is basic; rationale is not thoroughly explained or justified.

**Code Quality:** Code lacks organisation, and documentation is minimal or unclear. It is mostly working but there is a lot of redundancy.

**Analysis Depth:** Analysis is superficial, lacking in-depth exploration or meaningful insights.

**Visualisation Quality:** Visualisations are rudimentary and may not effectively convey insights or may be somewhat misleading.

**Storytelling:** The data story lacks cohesion, engagement, and a clear narrative structure.

**Originality and Creativity:** Very little to no evidence of originality and creativity, with heavy reliance on basic or generic approaches.

### Fail

**Dataset Complexity:** Dataset chosen is overly simplistic, or rationale is poorly justified or missing.

**Code Quality:** Code is disorganised, lacks documentation, and does not follow best practices. It might be mostly working but there are inconsistencies or there is not enough code to showcase that the student has mastered basic Python coding skills.

**Analysis Depth:** Analysis lacks depth, with missing or incorrect insights, and does not demonstrate a sufficient understanding of the data.

**Visualisation Quality:** Visualisations are missing, incorrect, or highly ineffective in conveying insights.

**Storytelling:** The data story is missing, incoherent, or fails to create a meaningful narrative.

**Originality and Creativity:** No evidence of originality or creativity; approaches used are incorrect or overly simplistic.

## Academic Integrity (Read this section carefully)

This is an individual assignment, and so all submitted work must be fully your own work. You are free to look up any online resources during your work on the assignment, however any attempt to communicate details of your solutions or approaches with others, whether other students or people outwith the University, is not acceptable. The use of any service, digital, online, or otherwise, to produce any submitted work in part or in full is not allowed. This includes any Large Language Model services or software like ChatGPT or equivalent.

The University of Stirling is committed to protecting the quality and standards of its awards. Consequently, the University seeks to promote and nurture academic integrity, support staff academic integrity, and support students to understand and develop good academic skills that facilitate academic integrity.

In addition, the University deals decisively with all forms of Academic Misconduct.

Where a student does not act with academic integrity, their work or behaviour may demonstrate Poor Academic Practice or it may represent Academic Misconduct, and the consequences for the student can be severe.

### Poor Academic Practice

Poor Academic Practice is defined as: "The submission of any type of assessment with a lack of referencing or inadequate referencing which does not effectively acknowledge the origin of words, ideas, images, tables, diagrams, maps, code, sound and any other sources used in the assessment."

### Academic Misconduct

Academic Misconduct is defined as: *"any act or attempted act that does not demonstrate academic integrity and that may result in creating an unfair academic advantage for you or another person, or an academic disadvantage for any other member or member of the academic community."*

The University recognises, amongst others, the following forms of academic misconduct:

* **Plagiarism:** *"a specific form of cheating which usually occurs when a student is working independently on an assessment (e.g. essays, reports, presentations or dissertations). Examples of other people’s ‘work’ can include anything taken from any form of publications, internet sources, the spoken word, graphics, data and written text. "*
* **Self-plagiarism:** where *"Duplicate submission of an item of written work in any other circumstance is not allowed and constitutes academic misconduct."* unless expressly allowed/required.
* **Inappropriate use of proof-reading:** *"Students cannot ask other people to write their work for them and should not use software (e.g., 'spinning' web sites, re-wording web sites or translation software) to generate text for them."*
* **Collusion:** *"When a student copies the work of another student either with or without the knowledge of the original author; or when two or more students work together to produce individual assessments."* This also applies to those that share their work with other students, even if they then do not copy the work and regardless of the intent.
* **Contract Cheating:** *"takes place when a student submits work for assessment that was completed by a third-party either for payment or for free. It is a broad category that includes, but is not limited to, work bought from so-called essay mills, customised work commissioned from ghost writers, and selling or exchanging work for use by others. A further example would be a friend or family member completing an assessment for a student. Work in this category covers the whole spectrum of assessment types. Any form of contract cheating constitutes academic misconduct, often of the most serious form."* More details about contract cheating can be found here: [Contract Cheating - Don't take the risk](https://canvas.stir.ac.uk/courses/694/pages/contract-cheating-dont-take-the-risk)
* **Dishonest Practice:** *"includes a wide variety of activities that aim to obtain an unfair advantage through (note this is not an exhaustive list):*
  + *Making false declarations to Faculties, Academic Staff members, Boards of Examiners or Appeal Panels.*
  + *Attempts to circumvent the similarity checking programmes that the University uses.*
  + *Submitting documents which have been forged in any way.*
  + *Attempting to gain or gaining access to examination or class test papers prior to their release and/or sharing examination or class test papers prior to their release.*
  + *Deliberate avoidance or refusal to engage with the relevant ethics review and approval process."*

The University of Stirling's full policy on Academic Integrity can be found at:

<https://www.stir.ac.uk/media/stirling/services/academic-registry/documents/policy-procedure-academic-integrity-v5-2023.docx>

# Assignment Instructions

## Exploring and Visualizing a Custom Dataset: A Data Story

The objective is to explore, analyse, and visualise a dataset of your choice, creating a compelling and professionally looking data story in a Jupyter Notebook. Your story should be supported by data manipulations and visualizations, and it should communicate interesting insights about the dataset.

You are provided with a Jupyter Notebook template which contains the 6 (six) parts you must address for the assignment. Each part is described below.

### Part 1: Dataset Selection and Justification

Choose a dataset that interests you. It could be related to sports, the environment, social media, etc. Ensure the dataset is rich enough for analysis (preferably having multiple columns and a substantial number of rows).

Write a markdown section explaining why you chose this dataset and what you hope to explore or find out.

### Part 2: Data Cleaning and Transformation

Load the dataset into a Pandas DataFrame.

Perform necessary data cleaning and transformation processes, explaining each step in a markdown cells.

Ensure the data is ready for analysis and visualisation.

### Part 3: Exploratory Data Analysis (EDA)

Conduct exploratory data analysis to understand the distribution and relationships within the data.

Use various techniques like calculating summary statistics, creating correlation matrices, and more.

Document your findings with appropriate markdown explanations.

### Part 4: Data Visualisation

Create a variety of visualisations using libraries such as Matplotlib or Seaborn to represent different aspects and insights from the data.

Each visualisation should be accompanied by a markdown cell explaining the choice of visualisation, what it represents, and any interesting observations.

### Part 5: Data Storytelling

Combine your analyses and visualisations to craft a data story. This story should be a cohesive narrative that provides insights into the dataset.

Your story should be engaging, making use of the markdown capabilities for formatting and structuring your content.

### Part 6: Challenges and Reflection

Discuss any challenges you faced during the analysis and how you overcame them.

Reflect on what you have learned from this assignment and how it relates to real-world data science tasks.

## Evaluation Criteria:

**Dataset Complexity:** Is the dataset rich and complex enough for a meaningful analysis?

**Code Quality:** Is the code well-organised and documented? Does it showcase Python scripting abilities or just the ability to call a few Pandas functions?

**Analysis Depth:** Does the analysis go beyond surface-level observations?

**Visualisation Quality:** Are the visualisations well-chosen and effectively communicate the data’s story?

**Storytelling:** Is the data story engaging, well-structured, and supported by the analysis and visualisations?

**Originality and Creativity:** Does the work show original thought and creativity in approach and interpretation?

### Notes:

* Ensure that the dataset you choose is **publicly available and cite the source**.
* You are encouraged to choose a dataset that has not been extensively analysed online to maintain the assignment's uniqueness.