A black background with white text

Description automatically generated with low confidence

MSc Data Science

Individual Project Module

7PAM2002-0509-2023

Department of Physics, Astronomy and Mathematics

**MSc Data Science Project Handbook**

**2023-24**

**Semester C**

**MSc Data Science Project Module Leader**

**Carolyn Devereux**

[**c.devereux@herts.ac.uk**](mailto:c.devereux@herts.ac.uk)

**University of Hertfordshire**

**College Lane, Hatfield,**

**Herts AL10 9AB, United Kingdom**

**Contents**

#### Introduction to the Project Module

#### Module Aims

#### Passing the module and where to get support

1. **Passing the module and late submissions**
2. **Where to go for support and contact details**

#### Choosing your project topic and dataset

#### Ethical issues and risk assessment

#### Supervision

#### Managing your project and practical work

#### Literature Review and Referencing

#### Academic Misconduct and Plagiarism

#### Assessments guidance

#### Summary of assessments

#### Presentation of Project and Data Management Plan

#### Ethics Quiz

#### Final Project Report (FPR) and Logbook

#### Viva

**ALWAYS maintain back-up copies of the most recent version of your code and final report so that they are in MORE THAN ONE location. Save your code in your OneDrive as well as at least one other place.**

1. **Introduction to the Project Module**

Welcome to an important part of your Master’s degree - your project. Your project is an opportunity to use the knowledge and skills you have developed during your studies so far and apply them to a substantial piece of independent work. Your project will be challenging, it will give you research experience and you will end up with a report that you can show to future employers.

The project is a showpiece opportunity for you to demonstrate what you know about current research and practices in data science and to demonstrate your skills in selecting and using appropriate techniques and tools employed to conduct a practical investigation into a particular problem. It is necessary for the award of a Master’s level qualification to demonstrate your ability to bring together a variety of skills, experience and knowledge derived from different sources.

Your project needs to include some practical work using data. You should answer a relevant research question and show that you understand the work you are doing as well as state the reasons for the choices you make to answer the research question.

The project is a self-directed piece of work that is conducted with minimum supervision. It demonstrates your ability to plan and manage a substantial piece of work and that you can direct your own efforts. You are expected to be thorough in your work and, in particular, identify and tackle any difficult or challenging aspects of the problem you are trying to solve.

Once you have started your project your supervisor becomes your personal tutor for the remainder of your course. Your supervisor should be your first point of contact for any questions or problems you may have. You can also contact the module leader.

This handbook provides guidance to MSc Data Science students undertaking their individual project. The successful completion of the MSc Project is worth 60 credits. You have to pass your viva to pass the module.

We hope you enjoy your project and the Data Science team look forward to working with you.

|  |  |
| --- | --- |
| MSc Data Science  Project Module Leader  Carolyn Devereux  [c.devereux@herts.ac.uk](mailto:c.devereux@herts.ac.uk)  Room 2E72A |  |

1. **Module Aims and Key Dates**

Successful students will typically:

* **Demonstrate knowledge and understanding of a specialised research topic or specialised application in data and relevant methods.**
* Be able to organise, plan and conduct a substantial research or investigative project.
* Be able to structure and produce a large research report.
* Be able to demonstrate understanding and debate/argue points of research interest in a viva style setting.
* Be able to present and debate a piece of on-going research or specialist application in data science.
* Be able to create and present a seminar talk on a specialised piece of research or specialised application in data science.
* Be able to review existing relevant literature and work related to a specialised research topic or specialised application in data science.
* Be able to follow good ethical practice (including avoiding collusion and plagiarism) when working.
* **Demonstrate knowledge and understanding of fundamental legal and ethical considerations in data science.**

1. **How to pass and where to get help**

**3a. Passing the module and late submissions**

**MSc Project entrance requirement**

To enrol on the project module, students must typically have studied 120 credits of taught modules from the MSc Data Science programme and passed at least 60 credits of taught modules (in order to satisfy UPR AS14, D5.2.3).

**Passing the module**

To pass the module you must get a 50% grade overall and pass your viva.

**Late submissions**

For each day (or part thereof) for up to five days after the assignment deadline, the late submitted coursework will have the numeric grade reduced by 10 grade points until, or unless, the grade reaches or is 50%. Coursework submitted later than five days after the deadline will be awarded a grade of zero (0%).

**Short-term Extensions and Exceptional Circumstances**

The Short-Term Extension (STE) for assessments (excluding the ethics quiz) can be granted for up to a maximum of seven (7) calendar days. Extension requests must be submitted prior to the assignment due date to the SASH team. [Applying for Short-term Extension](https://herts.instructure.com/courses/100562/pages/dot-dot-dot-about-short-term-extensions?module_item_id=2882060)

If you have good reasons for a late submission, not submitting, or failing to attend the viva or poster presentation, then you must complete the Exceptional Circumstances (EC) form with documentary evidence of the circumstances. This should be done as soon as possible and within 15 days of the assessment due date or in any case before the meeting of the Board of Examiners who will consider the matter. The form and process for applying for Exceptional Circumstances can be found here: [Applying for Exceptional Circumstances](https://ask.herts.ac.uk/exceptional-circumstances)

**What happens if you fail the module**

If you fail the module there are various outcomes and options:

1. If you get less than 20% in the module then you cannot have a referral and you will have to retake the module as a repeat module (FREN). You will take the repeat module in the following semester. Repeat modules have to be paid for.
2. If you fail the module and get between 20% and less than 50% then you can resubmit the failed assignments as a referral (FREF). The referrals take place during the next semester and the results of the referral will be available at the end of that semester. Referral assignments are capped to the pass mark of 50%. If you fail the referral then you have to repeat the module in the following semester and this will have to be paid for.
3. A deferral (FDEF) is allowed if you have applied for Exceptional Circumstances and it has been approved by the UH panel. You can then submit the Final Project Report and do the viva as a deferral in the following semester. Deferral assignments are not capped. Exceptional Circumstances must be applied for within 15 working days of the due date of the assessment.

**3b: Where to go for support and contact details**

It is important that you read and use the information in the Library SkillUp canvas site. It is particularly important to read and understand the referencing section.

[Library SkillUp and Referencing](https://herts.instructure.com/courses/113439/pages/library-skillup-how-to-research-and-reference-your-project?module_item_id=3548277)

**Library SkillUp**: It is strongly recommended that all project students use the modules on the Library SkillUp platform. There you will find a wide range of learning resources available for you and to help you through various stages of the project work. Each section is a short module that will help you understand the requirements for a research project. The modules are:

1. Searching
   1. This module will help you to do your literature review.
2. Referencing
   1. This module is essential to get the correct format for your references.
3. Writing
   1. This module helps you with your technical and academic writing.
4. Reading
   1. Learn to skim read many references before deep diving into a few papers.
5. Evaluating
   1. Learn how to evaluate your work professionally.

**SASH (SPECS Academic Support Hub):** Visit the [SASH canvas site](https://herts.instructure.com/courses/100562) where there is help with academic skills, academic English, technical writing, academic integrity and plagiarism and a list of databases for computer science. The SASH team also hold drop-in sessions or you can arrange a support session.

**Your Supervisor:** If you have a question or a problem while doing your project the first person to contact is your supervisor. For general module queries you can contact the Project Module Leader.

**General help:** Please make sure that the university has your most up-to-date email address and that you receive notifications of announcements .

|  |  |  |
| --- | --- | --- |
| **Team Name** | **Role** | **Email Address** |
| **SASH** | Academic Support | [SASH@herts.ac.uk](mailto:SASH@herts.ac.uk) |
| **UH Student Administration Services** | General student queries | [ask@herts.ac.uk](mailto:ask@herts.ac.uk) |
| **UH Student Immigration Team** | Student queries on visa issues and work | [sitadvice@herts.ac.uk](mailto:sitadvice@herts.ac.uk) |

1. **Choosing your project topic and dataset**

At the start of the project complete a ’Choosing your Project’ form which can be found in the Unit “Choosing Your Project’ on the Project module canvas site.

The project you choose must be related to data science and perform some type of analysis of data. The dataset could be a new dataset that has not been analysed before, or it could be an old dataset that has been looked at by other people but you wish to use for model development or to compare the performance of different models.

You should identify a research question that your project will answer. This can often be more difficult than it sounds so you should discuss your research question with your supervisor and continually review it throughout the project. You should clearly state your research question in your Final Project Report.

You can propose your own project topic or you can undertake one of the projects offered by the supervisors (the list of offered projects in the unit on ‘Choosing your Project’). You are working on this project for 4 months, and are expected to dedicate 600 hours of work to the project, so chose something that you will enjoy working on. It could be a topic in data science that you would like to explore in more depth, or it could be a dataset that you are interested in analysing, or you could pick an application that is based on an interest you have such as a hobby, sport or activity that you enjoy. A project is also an opportunity for you to learn a new topic in data science, for example, if you are new to time-series analysis or Natural Language Processing (NLP).

The type of activities that you can undertake in your project are:

1. Analyse a new dataset from a real environment.
2. Use a test dataset to develop a model.
3. Compare the performance of different models or versions of the same model.
4. Work on a mathematical based project, for example, developing or comparing the statistical analysis of your dataset.

You can work with a company on your project. The company may have data that they wish you to explore, or there may be some model development that you can do that is relevant for the company. You may have a link to a company, perhaps when you were on placement, that you could use to develop a good project.

**What makes a good project**

* Have an appropriate dataset and understand your data
* Pre-process the data
* Use suitable models for the objectives and dataset
* Have an appropriate training routine that takes into account overfitting/underfitting
* Explore improvements and optimise each model/method used
* Produce appropriate metrics and plots
* The results are meaningful and answer your research question
* Evaluation of the project and draw appropriate conclusions
* Include domain knowledge of your application
* A critical analysis of relevant published papers that are used to inform the project

**Choosing a good dataset is the key to a good project**

At the start of the project, make sure you have identified a dataset to work on. The dataset has to be suitable for your research task. The dataset has to meet UH ethical requirements.

You can find your own dataset or use one provided by your supervisor in one of the offered projects. There are many websites that contain suitable datasets. It is important that your supervisor agrees the dataset that you use for the project. Some websites that contain datasets are:

Worldbank <https://data.worldbank.org/> ,

World-in-data <https://ourworldindata.org/> ,

Kaggle: <https://www.kaggle.com/datasets>

1. **Ethical Issues and Risk Assessment**

**Ethical Considerations**

If you are planning to use data about people in your project, you MUST apply to the UH Ethics Committee for approval. This must be done before running a survey, conducting interviews, or getting people to evaluate your system. This includes asking someone else (even a friend or family member) to try out your code, for example, if you create an application then you cannot ask someone else to use the application without ethical approval.

It is important that the dataset you use has been collected ethically and it is your responsibility to make sure you know how the data was collected. If the data is online then this information should be on the website.

If there is data about people in your dataset then the data must be anonymised and the people must have given permission for the data to be used. Your dataset must meet ethical requirements.

**Health & Safety Considerations**

Activities undertaken during the project must conform to the Health and Safety requirements of the University and the law (see Health and Safety Guidance on Canvas).

**Applying to the UH Ethics Committee**

To get approval from the UH Ethics Committee you need to complete some forms. You should apply a minimum of three weeks in advance although it can take months to get approval. You should email all the documents to your supervisor to check before submitting to the Ethics Approval.

Your application will be reviewed by the University’s Ethics Committee. They have many forms to look at, so you need to give them enough information to decide quickly. For example, do not just say “conduct a survey” or “interview some clients”, because you haven't told them enough to let them judge whether there might be a problem with what you propose to do. You should explain the purpose(s) of the survey/interviews (e.g. to determine functional requirements, or to obtain feedback on the quality of the user interface) and say what the questions will be about. Write a statement that explains in detail what kind of questions you will ask (it would be better, and speed up the process, if you could provide the actual questions). However, if the questions you propose to ask are noncontroversial, and you promise to keep the responses anonymous, this should be a formality, but not trivial, AND it takes some time. Allow plenty of time.

Read the [UH Ethics Committee](https://herts365.sharepoint.com/sites/UHResearch/SitePages/UH-Ethics.aspx) guidance notes and find the relevant links for your application

**Not conforming to UH ethics rules can result in failure of your project.**

1. **Supervision**

You will be allocated a supervisor based on your ‘Choosing your Project’ form. You must agree the project topic and the dataset with your supervisor.

Your supervision sessions will generally be in a group. Your supervisor will set the time and day of your supervision meetings. Your supervisor may be unavailable during the course of your project and your supervision sessions will be organised around these dates. The supervision meetings will take place on campus and you must attend all the arranged supervision sessions.

For supervision meetings to be beneficial, you will need to make some progress between sessions. You should be working on your project full time, about 40 hours a week, and so you are expected to have plenty of work to discuss at each supervision meeting.

The supervisor’s role is to advise and support you on the project. They will help you to manage your project and be an academic advisor. Your supervisor will not tell you what to do, or do any part of the project for you, and so they are not responsible for your mistakes. You cannot necessarily expect that your project supervisor will be an expert in your chosen field, but they will be experienced in producing academic research and preparing academic reports.

Once you have started your project then your supervisor becomes your personal tutor for the remainder of your MSc course. Your supervisor should be your first point of contact for any questions or problems you may have.

**It is a general observation that students who meet with their supervisors regularly are more likely to achieve higher results in their projects.**

1. **Managing your project and practical work**

This is your research project and you set the agenda for it. You are responsible for the management of your work and you need to manage the time you spend on your project including the time you spend with your supervisor. To produce a good project report you should be working on your project daily and average 40 hours of work a week (600 hours for the whole module).

You will produce a project plan as part of your Project and Data Management Plan (PDM) and you should use it to plan what to do each week. Update it throughout the project and be clear what you are aiming to do each week to progress your project. This will help to keep you focused on your work. You will have to work on tasks in parallel to get all the work done in time, for example, you will need to work on your literature review and do your practical work at the same time. You should expect to learn new things while working on your MSc Project so allow enough time to learn what you need to. Make sure you allocate plenty of time to work on each of the assessments before the due dates.

To pass your project you must produce some practical work which means coding. You need to show some results and demonstrate your code in your viva to pass your project. So start doing your practical work early and finish it early so that you have time towards the end of the project to spend on writing your report. Not completing your practical work in time could result in you failing the module. When you write your results, analysis and conclusions sections you may realise you need some extra results so leave yourself time to do extra coding to add more results.

You may reach a point when you are not sure what to do next. If this is the case talk to your supervisor and agree a plan to go forward. You may also find there are times when it feels like there is too much to do. Again, you can talk to your supervisor and agree what are the most important things to work on and what can wait.

**All your work and results must be included in your final submission of the Final Project Report. The code is NOT marked as part of the report so make sure that anything in your code that you want to be marked is included in the report.**

Your code is reviewed during the viva but this will not add marks to your report if you have not included the work in the report.

Make sure any work you have done that explores different parameters, hyperparameters, methods or optimisation of your results, are included in the report. If you have done the work but it is not in the report you will not get any marks for it.

1. **Literature Review and Referencing**

**Literature Review**

The literature review is an important part of your project and requires time and effort to find good papers and to write about them. Looking at the literature gives you background information about the topic you have chosen and the methods you are could use. It also informs you about the research that has already been done that is relevant to your project and you should use this information to direct your work. For example, if one particular model has been shown in the literature to work best for your type of application then that is a justified reason for you to choose that model as your starting point in the project.

It is essential that you use peer reviewed published papers in your literature review. These are papers that are in published journals or in published conference proceedings. If your reference list is mostly websites and student theses then you will lose marks. It is common in the computing industry for papers to be published within peer reviewed conference proceedings to give a fast turn-around time for the publication. Computing is a fast moving industry and research changes quickly.

**Searching for relevant papers**

To find out more about how to search online and about the format for referencing and citation read the library SkillUp modules on ‘Searching’ and ‘Referencing’.

[Library SkillUp and Referencing](https://herts.instructure.com/courses/113439/pages/library-skillup-how-to-research-and-reference-your-project?module_item_id=3548277)

**Use the UH online library.** You can search for published papers using any search engine but you may not be able to read some of the publications without paying. If you search for the publications using the UH online library then you will be able to read them since many of the publications that you will be interested in have been subscribed to by the library so that you have free access to the whole paper.

**Referencing**

**Referencing and citations must be in Harvard format alphabetically.** Read the [Guidance on referencing](https://herts.instructure.com/courses/113439/pages/library-skillup-how-to-research-and-reference-your-project?module_item_id=3548277) .

* The reference list at the end of your report must be in alphabetical order based on the first author surname.
* At the end of each reference you must include a linked web-address (see format below).
* All citations in the report must be in the reference list.
* The reference list must only contain the publications that are cited in the report.
* All papers referenced should include the author’s list, year of publication, the journal name, the journal volume and the page numbers.
* If the paper is from arXiv include the authors, the year, the publication number and the web address.
* Any websites quoted should have an author name (or company name) of the publisher and the year it was published (both are often at the bottom of the page in very small print and you use both for your citations) and then the web address. You must also state when you accessed the website.
* If the publication is a book then you include the author name(s), the year, the book title and the publisher.
* If you want to include books, websites or other information that you read as background that you think is relevant but is not a specific reference then you have to include these in a ‘Bibliography’.

**All references must be real references that can be accessed by the marker. A fake or incorrect reference is an academic offence.**

**In-text Citations**

A citation are the words put in the text of the report that show where the reference relates to. Citations should be put at the first place the author or publication is mentioned.Citations should not be put at the end of the sentence of paragraph if you first mention it at the beginning or middle. Citations come in two forms;

1. ‘First author surname (date)’ - for those where the author’s name is part of the sentence

**Example:** “Smith et al, (2020) said that …”.

1. ‘(first author surname, date)’ - for those where the author’s name is not mentioned in the text

**Example:** “It was shown (Smith et al, 2020) that…”.

**Examples of referencing**

**Book**

Author, initials (year) ‘Title of book’, Publisher. (Available at URL)

**Example**

* Frank, R.H., (1997), *Microeconomics and behaviour*, London: McGraw-Hill (Available at: <https://www.mheducation.co.uk/microeconomics-and-behaviour-3e-9781526847843-emea-group> )

**Journal publications**

Author(s), Initials, (Year), ‘Title of the article’, Name of the journal, Volume number, (Part number), First and last pages. (Available at URL)

**Example**

* Watson, M., (2006), ‘Management accounting and budgetary control’, *Public Finance Quarterly*, 3 2(2), pp. 234-7. (Available at: <http://search.global.epnet.com> )

**Website**

Author or company name, (Year), *Title of website or* *Name of report,* [Online]. (Available at URL), [Date accessed]

**Examples**

* Mintel (2006) *Holidays on the Internet* [Online] (Available at <http://reports.mintel.com/> ) [Accessed 8th June 2006]
* National Statistics (2006) *UK Trade* (Available at: <http://www.statistics.gov.uk/cci/nugget.asp?id=199> ) [Accessed 8th June 2006]

1. **Academic Misconduct and Plagiarism**

It is important that you read and understand the following pages:

[AskHerts Assessment Offences and Academic Misconduct](https://ask.herts.ac.uk/assessment-offences-and-academic-misconduct)

[SASH page on Academic Misconduct](https://herts.instructure.com/courses/100562/pages/understanding-academic-integrity-or-how-to-avoid-unexpected-problems-with-your-assessments?module_item_id=2573770)

[SASH page on Plagiarism](https://herts.instructure.com/courses/100562/pages/plagiarism?module_item_id=2578534)

Academic Misconduct is any action which gains, attempts to gain, or assists others in gaining or attempting to gain unfair academic advantage. Academic Misconduct comes in various forms which include plagiarism, collusion, contract cheating, fabrication of data, copying text, and false referencing. This also includes the use of essay mills and support sites which have been criminalised in the UK.

If it has been reported that you may have committed an assessment offence and/or academic misconduct, an independent person within your School will investigate the allegation made against you. In the first instance, Academic Misconduct offences will be dealt with in accordance with the procedure set out in this document ([Appendix III, UPR AS14](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct.pdf)).

**If academic misconduct is suspected you risk failing your module or being withdrawn from the University.**

**Using Turnitin**

Your submissions will be checked for plagiarism using the software ‘Turnitin’. When you submit your assessment the documents will be checked by Turnitin and you can see the overall score of your submission on canvas. It is expected that the references, key words and phrases, the front page and the declaration page will be picked up as similar by Turnitin so a score of 10% to 20% is normal. Markers will be looking for similarity of paragraphs or multiple sentences that show that they have been copied from another source. If copying is evident then these sections will not be marked which could result in failure of the module and you could be reported for academic misconduct.

**Plagiarism**

The use of other people’s work under one’s own name is a form of academic dishonesty that is illegal and unacceptable to the University. It involves copying work produced by other individuals without acknowledgement to them and using text that has been generated by AI (including chatGPT). You must ensure that you always acknowledge material taken from books, journals, company information etc.; you can do this by using quotation marks for direct quotes and referencing any relevant sources. Correct referencing is an important part of not plagiarising.

**ChatGPT**

It is an academic offence to use generative AI (including chatGPT) to write any part of your report or code without declaring it and you risk getting a zero grade for your report. If you use chatGPT as part of your project then you must declare this and reference any parts of the report and code where it has been used.

**Collusion and Cheating**

It is an academic offence to submit a report that is written by someone else or by generative AI. It is also an offence if the computing code you submit is not written by you.

Collusion is an academic offence. Collusion is when:

* you present work that has been done jointly with another person as if it is your own work completed independently.
* there are strong similarities between the work of two or more students in an individual piece of coursework.
* you lend your work to someone and they copy it.

If you are found to be colluding you may both fail the module and be reported for academic misconduct.

**Viva**

As part of your viva you need to present your report and code to show that you have written both the report and the code. You will be expected to answer questions to show that you understand the content.

**If your work is found to include plagiarism, collusion, cheating or other forms of academic misconduct then severe penalties could be imposed and you risk failing the project module or failing your MSc course.**

1. **Assessments (Guidance and Marking Criteria)**

**Summary of Assessments**

1. **Presentation of Project and Data Management Plan (10% on campus)**
2. **Ethics Test (10% on campus)**
3. **Final Project Report (FPR) and Logbook (70% submission on Canvas)**
4. **Viva (10%, on campus)**

**10i. Presentation of Project and Data Management Plan (PDM) (10% of module)**

**Marking Criteria – PDM Plan**

|  |  |
| --- | --- |
| **PDM Plan (60%)**   * Quality of research question and difficulty of project * Suitability of plan to meet objectives and assessments * Evidence that the data meets ethical requirements * Suitable plan for collecting, storing and version control of data * Relevance of literature identified and correct format of references | **Presentation (40%)**   * Quality of presentation * Response to questions |

**The Presentation**

Your presentation will be to your supervisor and supervision group. Your supervisor will arrange the day and time and it will be on campus. During the session you will present your DMP plan and listen to the other presentations in the group. Your presentation will be a maximum of 10 minutes followed by questions up to a maximum time of 15 minutes. Prepare for your presentation by practising (and timing) your presentation. You must give a presentation to get any marks for this assessment. If you miss the presentation and submit your plan then you will still get a zero (0%) grade for the assessment.

**The PDM Plan**

The Project and Data Management Plan is a summary of your project topic, research question, project plan and data management plan. The requirements for the PDM plan can be found in the assignment section on canvas.

The PDM plan should include the following:

1. **Project Title.** Make your title specific to your project and descriptive about what you are doing in the project.
2. **Research Question.** Your projectmust include a research question that you will answer through your project. The research question must allow you to produce an analysis of an issue or problem. It must be about what your project is doing and not what the work could be used for in future. The research question can be stated as a hypothesis statement instead which is a statement that will be proved or disproved by the project.

A good research question narrows the topic of interest to be a specific area of study and will help identify the appropriate methods, data, and data analysis. If the research question is too broad it will not be helpful in determining your aims and objectives. ([Lipowski, 2008](https://academic.oup.com/ajhp/article-abstract/65/17/1667/5128061?login=false) ). You must agree your research question with your supervisor.

1. **Aims and Objectives**. The aims and objectives are the methods by which you will answer the research question. Include a summary of the approach you are planning to use to answer your research question, for example, are you going to compare 3 machine learning models or analyse the dataset using clustering and visualisations. Which models will you use? You need to be specific in the plan although you can change your methods and analysis during the project.
2. **Background.** A paragraph on the background to the topic (approximately half a page). This will be an overview of your project idea and include some background on the topic including a few citations of at least 2 or 3 relevantpapers.

1. **The dataset** that you will be using as well including the source of the dataset with an appropriate reference. A brief description about what the data includes and size of the sample.
2. **Data Ethics.** Statement on ethical considerations for your project plan. Consider the dataset being used, what you will be doing with the data and how you will test the code. Any involvement of people, other than yourself, are likely to require approval by the University of Hertfordshire Ethics Committee. Any data that uses personal data must be anonymised. If there are no ethical issues then state why none apply.
3. **Project plan**. Break your project into key tasks. Write a short sentence about what each task will do. Prepare a simple Gantt chart (a timeline) that shows your key tasks and key milestones. Make sure you include all the assessments with the due dates. A simple excel spreadsheet on a two-weekly basis is suitable. You should make sure that you are multi-tasking meaning that your tasks should not be one after another, for example, you should be doing your literature review at the same time as your practical work.
4. **Data Management Plan.** Summarise how you will collecting the data, store the data and code, backup the data and code, and the version control you will be using. You should use GitHub to track the development of your code.
5. **Reference list and bibliography.** References must be in Harvard format in alphabetical order (see section 8). They should be mostly published papers and it is acceptable to have a few websites in your reference list but you will lose marks if you have mostly websites. A bibliography is a list of resources you have used to research your idea; this can include books and websites that you have used for background reading.

**10ii. Ethics Test (10% of module)**

This assessment will be an online test that will cover the issues of ethics in data science, the laws on storing and using data, ethical research, referencing, and academic misconduct. The questions will include the ethical issues, risks and requirements under the UH policy.

**The questions in the ethics test will be based on the content covered in the project lectures and reading list on canvas. It is essential that you attend the lectures.** The questions will be multiple choice. The test will be on a specific day and time and will take place on campus. You must attend the test at the allocated day and time.

**10iii. Final Project Report (FPR) (70% of module)**

**For this assessment you must submit the Final Project Report and the Logbook.**

**Make sure that you submit your report and logbook well before the due date/time. Being late will lose you 10% of your grade for each day late. You have to do a lot of work on your report to gain 10% so it is not worth losing 10% for being late.**

**Logbook (5%)**

**Marking Criteria - logbook**

Your logbook grade will be based on the following criteria:

* Attendance and participation at supervision and lectures.
* Versions of the code on GitHub showing the development of the code.
* Well written user documentation of the code on GitHub.
* Good data management.

**All versions of your code must be included on a GitHub site.**

You must submit a logbook of your activities when you submit the Final Project Report. The logbook is submitted as a separate document. The template for the logbook is on the module canvas site. The logbook must record your attendance at lectures and supervision, your weekly activities, your code development (code versions on GitHub), and the link to your GitHub site. You must make sure that the markers have access to your GitHub containing all the versions of your code.

**User documentation for the code must be included in your GitHub site.** This is a description of what the code does, format of the input data required, format of the output data/plots/tables, summary of parameters and hyperparameters used, and any relevant issues that you had developing the code.

**Final Project Report (65%)**

**Marking Criteria - FPR**

Your Final Project Report grade is based on the academic judgement of two markers and will be based on the following criteria:

* Clear project specifications:
  + Abstract, introduction, objectives, ethical considerations, evidence that the objectives were met, difficulty of project.
* Quality of research and literature review:
  + Critical assessment of relevant published papers, depth of review, clear description of the background to the project, application and methods.
* Evidence of good practical work:
  + Suitable data and pre-processing, appropriate choice and justification of methods (based on literature and type of data), good implementation of methods and code, novelty of methods and code, evidence of training and testing improvements/optimisation including parameter and hyperparameter trials.
* Evaluation of results and justified conclusions:
  + Quality of results, suitable metrics used, in depth analysis relating the results to the literature and the objectives, future work identified.
* Presentation of report:
  + Use of the correct front page and declaration page, quality of technical writing (including language and grammar), structure of the report and professional layout, appropriate referencing and citations (including correct format), figures and tables correctly labelled and credited, within the word count.

**The Final Project Report**

**Word count and format**

The Final Project Report should be a maximum of 7,500 words, this means that the report can be less words but cannot be more. The word count includes the abstract and contents page and the subsequent sections up to and including the conclusions. The word count does not include the reference list, the appendices, the front page, the declaration page, and the acknowledgements.

The report must be written in either Arial, Times or Calibri font with a font size of 12 and single line spacing. The Final Project Report should include sections that are relevant to your project, if you are unsure about the sections to include then talk to your supervisor. The following sections give an outline of what should be included.

**Front Page and Declaration Page:** **You must use the template provided in the Assignments section on canvas for the front page and declaration page.** **You must sign the declaration page.** You must add the word count on the front page. The blue writing on the front page template should be replaced with the information on your project.

**Acknowledgements:** Include acknowledgements if you wish. This is purely your personal choice and you can choose who you wish to mention if you add an acknowledgement. There may be some situations where you have been asked to include acknowledgements, for example, if you have used company data or certain software packages that ask you to include them in the acknowledgements. Talk to your supervisor about what to include if you think this applies to you.

**Abstract:** This should be a summary of your whole report: your research question and objectives, your methodology and dataset you are using, your results and analysis, and your conclusions. It should be one paragraph only with no references included. The abstract should be before your content table and it does not have a section number.

**Contents page:** Include a contents page that is 1 to 2 pages long. Do NOT add a list of figures or list of tables.

**Introduction:** This should give an overview of the purpose of the project and the application it relates to and, if relevant, say what is currently being done in the industry. The research question, aims and objectives of the project should be clearly stated (you may want to have them as a sub-section).

**Background:** This includes your literature review with a suitable number of references with correct in-text citations. Give a clear overview of the technical background to the project; this should be computing based since this is a data science project. It is important that you demonstrate some in-depth critical analysis of four (4) or more relevant published papers in peer reviewed journals or conferences (not websites or from a thesis). A table containing lots of papers is not a good literature review unless you have also included a more detailed critical analysis of individual papers.

Start your literature review with an introduction to the literature on the subject and why you chose the papers that you did (what was your selection criteria). Then discuss in detail some of the key papers that are relevant to your project as a critical analysis. It should be clear why you are discussing these papers. **A critical analysis includes:**

* what work was done,
* what data was used,
* what methods were used,
* what were their results and conclusions,
* how the paper relates to your project,
* what is good and what is limited about their work.

**Dataset:** Include a section that describes your dataset. State where you got it from including a full reference to the exact website. Describe how the data was collected originally (not how you got it but how was the data made/put together), who collected it, which country and when was it collected, why was it collected, and what the data includes (this is required even if you got the data from a website). State why you chose this dataset to answer your research question and justify the reasons.

You should always look at your data before starting work on it so this section should include your Exploratory Data Analysis (EDA) that shows relevant images/tables/plots of the data. Discuss any data pre-processing you have done. Be specific and detailed about the work you did. For example, if you removed null data then, what format was the null data in before and after, how many records were affected, what was the impact on your results of changing the null data? Only discuss the pre-processing that you have done.

**Ethical Issues:** You must have a statement about the ethics of using your dataset. Even if there were no ethical issues identified you have to show that you thought about the issues to see if there were any relevant to your project. Include the ethical issues that you have considered regarding the dataset and project along with any evidence.Ethical things to consider (including if you got the data from a website):

* Is personal data included and if so is it anonymised?
* Does your data come under GDPR?
* Does using your data require UH ethical approval? For example, do you collect personal data from the internet (e.g. data from social media)?
* Does your project require UH ethical approval? For example, do you collect data from people or do a survey?
* Do you have permission to use the data? Is there any evidence that you can use it, for example, a creative commons licence, and if so then include a screenshot of it in your report? Do you have to pay to use the data?
* Was the data collected ethically? This can be the most difficult thing to determine. For example, if the data includes personal data then did the participants give their consent for the data to be on the website and be used for general research? If your data is from a website then you must explore the whole site to find out as much as you can about the data. Is there a reference to a published paper on the website that states how the data was collected, if so this should be discussed in your report? Was the data collected (or put on the website) by a reputable organisation so that you can assume it was collected ethically? Did the data come from another site that contains the original data in which case you should use the original data (if you do not use the original data then state why you did not)?

**Methodology:** This is likely to be a long section. This describes the practical work you have done. It has to be a specific description of the technical work that you have done in your project. What did you do and how did you code it? It is likely that you use the words ‘I’, ‘me’ and ‘my’ a lot in this section to show that you did this work. It needs to be as technical as you can. State the exact models you used and why you chose them. State the exact metrics you used and why you chose them. If you include techniques that you did not use then you will lose marks.

**Results:** Think carefully about which metrics you use and make sure you know what you are measuring and why for your project. For example, what is accuracy measuring for your project? In general, you should be using more than one metric. Make sure your metrics meet the project objectives. Things to think about when producing your results:

* What are the best metrics to use? Do not use just one metric - different models may perform better when measured in a different way. Consider the appropriate mathematics behind a metric and whether that suits your data type, model and research question.
* What is the best way to present your results? For example, if you are comparing models then a single plot or table comparing all three models could be a good way to see the comparison. Should you include a confusion matrix? Are your results better presented in a table or multiple plots? If your data is images then show plenty of examples of the images.
* Understand and write what each metric/result means for your project and what you are measuring.
* What do the results mean for your application or for other applications? Can your project be used in a real world situation?
* Do the results address your research question?

**Analysis and discussion:** The analysis section is what can turn a good report into an excellent report. Consider the following issues:

* What do your results mean?
* Which model works the best and why? What is it about the way a model or method works that makes it work well or poorly with your data?
* How do the results compare to the literature that you have discussed in your background section? Why do you think your results are better or worse than the literature?
* What are the limitations of your results?
* How do the results relate to the project objectives?
* How do the results relate to the project application/topic/research question?
* Are any of the models useable in a practical situation and if so why and how?
* Discuss whether you have answered your research question.

**Conclusion:** This should be a short section (between one paragraph and two pages). It should include:

* a summary of the key results,
* your justified conclusions,
* the applications and real world situations that your work can be used for,
* the future work you would recommend (what would you do next if you carried on with the project?).

**References:** This should be a full list of all the references that you cite in your report. All referencesshould have an in-text citation and all in-text citations should be in the references list. The references must include the peer reviewed journals that the papers were published in. **If most of your references are websites you will lose marks.** The references must be in Harvard format (author name and year). The reference list must be in alphabetical order based on the surname of the first author (not numbering). For the correct formatting of your reference list and in-text citations see Section 8 ‘Literature Review and Referencing’. You can shorten conference proceedings e.g. 4th Int. Conf. Comms. & Comp. Tech. You will lose marks for getting the formatting wrong for your citations or for your reference list. **It is an academic offence to include any fake references.**

**Appendices: You must include your code in text format as an appendix.**

The appendices provide supporting evidence of the quality and quantity of the work you have done.Include information that you think is relevant as an appendix. Discuss what to include in your appendices with your supervisor. Do NOT include a Gantt chart into the report (this was a task for your Project Plan). In your appendices include the following:

1. **Extra plots or images:** If you have a lot of plots or images in your results consider putting just the most important ones into the report (a small selection that show the main results) and the rest of them into an appendix.
2. **Code:** You must include the code that you have written and it must be in text format. Do not include your code as screenshots. Put your code as the last appendix since it can be quite long. Do not include code that is machine generated or that comes from a different author unless it is necessary for the reader to understand the work you have done. If you do include code that you did not write yourself, it is your responsibility to make clear which parts of the programme are your own and which parts are not (use referencing and comments). If you present automatically generated code, or the code of another programmer, as if it were your own then you are committing plagiarism.

**All your work and results must be included in your final submission of the Final Project Report. The code is NOT marked as part of the report so make sure that anything in your code that you want to be marked is included in the report.**

Your code is reviewed during the viva but this will not add marks to your report if you have not included the work in the report.

Make sure any work you have done that explores different parameters, hyperparameters, methods or optimisation of your results, are included in the report. If you have done the work but it is not in the report you will not get any marks for it.

**General comments about writing and submitting your report**

To turn any writing from adequate to excellent is about editing, editing, editing. Re-read and re-edit your report as much as possible. Make the writing flow so that you ‘tell the story’ of what you did in the project and the reader can follow the logic of your descriptions and arguments (i.e. they can follow what you did and why you did it). When you are writing your report make sure that you understand every sentence you write and can explain each paragraph then you will be ready for the viva.

**Make sure all figures and tables are numbered and labelled** stating in a short sentence what the figure or table shows. Make sure diagrams or photos do not need paying for under copyright. Each figure and table should be mentioned in the text of your report (except general photos/pictures). Any figures/photos/pictures/tables that are not your own (from a website, paper, report etc.) must have the source (credit) underneath the image stating where you got it from with full website or reference – this website/reference does not need to be in your reference list. Some diagrams request that you put an author or company credit on them. **If you do not include a credit stating where you got the image /table /figure and you did not produce it then that is plagiarism.** If you include a flowchart it should be one that you have drawn that is specific to your project. To make the flowchart specific add the exact models, the exact metrics, the specific pre-processing techniques you used etc.

If you need help with your academic English or academic writing then you can get help at the LRC in the support services and also at SASH (SPECS Academic Support Hub). Getting someone to proof-read your report could be counted as academic misconduct unless you state that it was proof-read and the name of who read it. Getting someone to write parts or all of your report is an academic misconduct that could result in you failing the module and possibly the course.

**ALWAYS maintain back-up copies of the most recent version of your code and final report so that they are in MORE THAN ONE location. Save your code in your OneDrive as well as at least one other place. Losing your code or report due to computer failure is not an acceptable reason for Exceptional Circumstances and could result in failing the module.**

**10iv. Viva (10% of module)**

**You must pass your viva to pass the project module. Make sure you can talk about everything in your report and code.**

**Marking Criteria - Viva**

Your Viva grade is based on the academic judgement of two markers and will be based on the following criteria:

* Clear summary of project and good understanding of project content and aims:
  + objectives, data, methods, results and analysis.
* Response to questions:
  + ability to answer questions (level of difficulty), clarity of answers, clear technical explanations, shows an understanding of the data science used in the project, shows an understanding of the wider implications of the project (society, technically, commercially).
* Demonstrate a good understanding of the project code.
* Professionalism and communication skills:
  + speaking clearly, good use of technical vocabulary and academic English, confidence, preparedness and being on time.

**The viva**

A viva is a technical interview where two UH academic staff will ask you questions about your project and code. The purpose of the viva is to demonstrate your knowledge of your project and the data science that you have used in your project. You are expected to demonstrate your code so it is essential that you have your code ready to show on a computer. You can be asked questions on any line of your code and you may be required to run specific sections if asked. If you have developed an app then you should demonstrate the app working.

**The viva will take about 20 minutes (with a maximum of 30 minutes). The viva will be on campus.** The date and time of your viva will be set by your supervisor. Your markers will normally be your supervisor and a second member of academic staff. You must attend promptly.

Spend the first 2-3 minutes summarising your project – focus on your results and your conclusions. **Do NOT prepare a powerpoint presentation.** For the remainder of the viva you will be asked questions about your report and code.

**A viva is an exam, if you are late or miss your timed session then it is the same as a missed exam. Only emergency situations are an excuse for missing a viva. If you do not attend your viva then you will receive a zero mark for the viva and you will fail the module.**

**Format of the viva**

1. At the start of the viva you will give a 2 to 3 minute summary of your project that includes an overview of the project aims, the data, the methods you used, the results and the conclusions. Make sure you can say how your work has met the objectives of the project.
2. You will be asked questions about your project and report.
3. You will be asked questions about your code and you will need to show your code. Your demonstration should show the output of the code, for example, if you have used a notebook then the code outputs should already be seen in the notebook. If you have developed an app then demonstrate the app working.
4. You will have an opportunity at the end to give any final comments.

**How to prepare for your viva**

Practise talking through a summary of your project and time how long it takes so that it is a maximum of 3 minutes. Practise talking through what each line of your code does. Think about what questions may be asked and practise answering them out loud or to another person.

* Write a list of possible questions and prepare bullet point answers against each one; this should include questions about your report and your code. What would you want to know about your project?
* Be able to explain all the technical terms you have included in the report.
* Be able to provide technical explanations of your models, pre-processing and methodology.
* Be able to explain all the results you have shown in your report – this is essential.
* Be able to discuss your analysis and conclusions and how your results could be used in practical applications and real world situations.
* Be able to talk about the code you used to produce the results (e.g. which tools and functions did you use, how many layers, what parameters etc.). You should be able say what you have done and why you chose to code the way you did.