



Faculty of Information Technology

Semester 1, 2025

FIT5145: Foundations of Data Science

Assignments 1 & 3: Business and Data Case Study

1. **Assignment 1 (Proposal):** Draft a proposal to introduce a data science project of interest. The due date of Assignment 1 is: **Friday, 4 April 2025, 11:55 PM (Week 5).**
 2. **Assignment 3 (Report+Presentation):** Write a comprehensive report on your data science project and prepare a 4-minute presentation on your project. The due date of Assignment 3 (the final project report and presentation slides) is: **Friday, 23 May 2025, 11:55 PM (Week 11)** and the presentation will be held **in the Week 12 applied class.**
- Both Assignment 1 and 3 are **individual** assignments.
 - **Please do NOT zip your submission files.** Zip file submission will have a penalty of 20% of the total mark of the assignment.

Focus of the Project Proposal

Assignments 1 and 3 require you to develop a **novel** data science project proposal that introduces an original approach to solving a significant real-world problem using data science methods. You are expected to go beyond existing studies by identifying unique problem statements, proposing innovative methodologies, or applying data science techniques in new contexts. Your proposal should demonstrate your ability to define a novel and important problem, identify relevant datasets, select appropriate methodologies, and develop effective evaluation strategies. The proposed project should align with the following business scenarios: **agriculture, education, finance, gaming industry, healthcare, social media, and sports**. You are encouraged to discuss any project ideas with your tutors for further guidance.

Assignment 1: Proposal (15%)

Weight: 15% of the unit mark

Submission format: one PDF file

Size: up to 1000 words.

What you need to do:

- Choose a data science project.
- Write the initial three sections: (1) *Introduction*; (2) *Related Work*; and (3) *Business Model (References* as well to support your project) of the report, as detailed in the specification of *Assignment 3: Report + Feedback + Presentation* below.

We have developed a system named FLoRA to support you to accomplish Assignment 1, which you may access via: <https://www.floraengine.org/moodle/my/courses.php>. You are expected to work with the

chatbot embedded in FLoRA, powered by cutting-edge Generative Artificial Intelligence (GenAI) technology GPT-4o, to identify a novel and important problem to tackle in your data science project. You can discuss the assignment requirements with the GenAI-powered chatbot, seek suggestions for potential project topics in a specific domain, and gather information relevant to a specific project topic by conversing with the chatbot. Even more, you may send the proposal draft to the chatbot and seek feedback for further improvement.

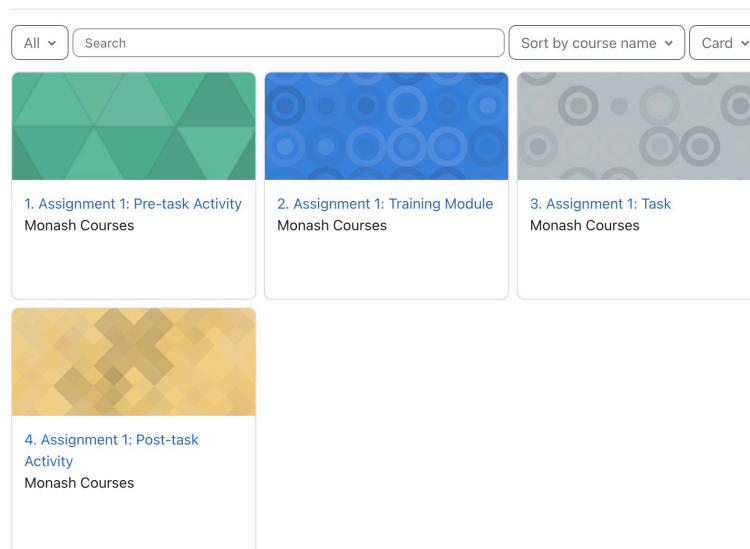
Please notice:

- **We will send you the login credentials via emails for you to access the FLoRA platform.**
- You are only expected to use FLoRA to accomplish Assignment 1, though you may use it for Assignment 3 as well, it is not mandatory.
- The conversational data you generate when interacting with the GenAI-powered chatbot will be used for textual analysis in **Assignment 4** in this unit. That is, the conversational data will be shared with the whole class. Therefore, please notice the following:
 - Please **only** discuss your data science project with the GenAI-powered chatbot **in English** and do not ask any questions that are irrelevant to your project;
 - Please do not disclose any personal or sensitive information when interacting with the GenAI-powered chatbot.
 - There are **four modules** in FLoRA and you are required to complete all of them. **Important:** *Any incomplete activities in these modules will result in your conversational data with the chatbot being excluded from the dataset used for Assignment 4, preventing you from answering the questions in Assignment 4.*

After logging to the platform, you will see there are four modules required to accomplish Assignment 1, as shown below:

My courses

Course overview



Module 1: Pre-task activity

- Please provide information about yourself as well as your prior knowledge and experience in data science and GenAI.

Module 2: Training Module

- We provide a set of tutorial documents to help you familiarise with FLoRA, including:

- *The system interface;*
- *The annotation tool*, which you can use to make annotation to the reading materials provide to help you get some inspirations about potential project ideas before discussing with the GenAI-empowered chatbot;
- *The essay writing tool*, which you can use to draft the assignment;
- *The GenAI-empowered chatbot*, which you can consult for help when solving the assignment. The Chatbot uses the GPT-4o model.
- Please notice that all these tools have been enabled in Module 2 (available on the top right corner) and you may familiarise yourself with these tools first before moving to Module 3 to start working on the assignment.

Module 3: Task - Assignment

- This is the main module in which you are expected to accomplish Assignment 1. Before discussing with the GenAI-powered chatbot, please first have a look at the “inspiring” materials that may give you some initial ideas of what data science can achieve in the domains where data science is playing an increasingly important role:
 - Data Science in Agriculture
 - Data Science in Education
 - Data Science in Finance
 - Data Science in Gaming Industry
 - Data Science in Healthcare
 - Data Science in Social Media
 - Data Science in Sports

You may use the provided annotation tool to annotate the useful information in these materials.

- After selecting the domain that you would like to work on, use the GenAI-powered chatbot to get necessary help for accomplishing the assignment (e.g., seeking relevant information about a specific topic in the selected domain).
- After you finish the draft, please (i) click the “Save Essay” button to send your submission to FLoRA; and (ii) copy your project text and paste it into a word processing tool (e.g., Microsoft Word), format it if necessary, and then save it as a PDF file and submit it on Moodle as well. **Important:** *Please ensure that the final project text saved/submitted in FLoRA matches the PDF version submitted on Moodle, as the FLoRA submission will be used for peer grading, as detailed later. You may save/submit the written report multiple times, and the final version saved/submitted will be exported for peer grading.*
- Please notice that the conversational data you generate with the GenAI-powered chatbot will be used and shared for Assignment 4 and thus do not disclose any personal or sensitive information to the chatbot.
- As we will use the conversational data for Assignment 4, ideally you have one “meaningful” discussion session with the GenAI-powered chatbot (instead of having multiple at different times) in Module 3 to get the help you need to accomplish Assignment 1. Prior to this, you may familiarise yourself with the chatbot (and other tools as well) in Module 2.

Module 4: Post-Task Activity

- Please share your experience in using FloRA as well as the GenAI-powered chatbot to tackle Assignment 1. Your responses to these survey questions will be **mandatory** for including your

conversational data to prepare an anonymised conversational dataset for Assignment 4.

For any technical issues in using FLoRA, please contact guanliang.chen@monash.edu and xinyu.li1@monash.edu

Assignment 3: Report (15%) + Feedback (5%) + Presentation (10%)

1. Assignment 3: Report

Weight: 15% of the unit mark

Submission format: one PDF file and one RMD file (for demonstration in the Characterising and Analysing Data section)

Size: up to 2500 words

This report is your *comprehensive* analysis of how data science can be used to help solve a significant real-world problem. Please answer the following question in the **FIRST** page of your Assignment 3 submission:

- Have you selected a topic for Assignment 3 that is different from the one that you used for Assignment 1 (i.e., have you rewrote the first three sections of the report)?

Your report should have the following sections:

1. Introduction

- Clear articulation of the specific problem the project aims to solve.
- Background and context of the problem.
- Importance of the problem (why it matters).
- Specific goals of the project.

2. Related Work

- Summary of existing research, projects, or industry solutions related to the problem.
- Identification of gaps in current approaches.
- Why or how your project should be considered as *novel*.

3. Business Model

- Analysis about the business/application area the project sits in.
- What kind of benefits or values the project can create for the specific business area?
- Who are the primary stakeholders and how will they benefit from the project?

4. Characterising and Analysing Data:

- Discuss potential data sources and analyze their characteristics (e.g., the 4 V's), evaluate the required platforms, software, and tools for data processing and storage based on the specific characteristics of the data or consider potential options (e.g., platforms, software, and tools) if your project expands in the future.
- Specify the data analysis techniques and statistical methods (e.g., decision tree or regression tree) applicable to the project. Provide a rationale for the selected methods and discuss the expected high-level outcomes. Note: The specification of data analysis and statistical methods should be different from the demonstration below and must be described separately.

- Demonstration: identify a usable dataset for the proposed project and perform some basic analysis on the identified dataset to demonstrate the feasibility of the project, using R (e.g., detailing the information/features contained in the dataset, analyse the basic characteristics of the dataset, etc.), and report the analysis process and result in the demonstration section of a final report.

Note: Please include a link to download the dataset in the final report, and upload the R markdown file created for data analysis on Moodle.

5. **Standard for Data Science Process, Data Governance and Management**

- Describe any standards used in your data science process
- Describe any practices for data governance and management in the project, e.g., how to address key issues such as data accessibility, security, and confidentiality, as well as potential ethical concerns related to data usage.

The sections would present aspects of Weeks 1-10 of the unit for your chosen case study.

The **maximum word limit** for the report (Assignment 3) is **2500** words. It may include some/all of your Assignment 1, modified if needed (counted in the 2500 word total). References at the end of the report (i.e., URLs and academic publications) are not included in the word count. Note that staying within the word limit demonstrates your ability to write concisely.

2. **Assignment 3: Feedback from Assignment 1**

Weight: 5% of the unit mark

Please ensure the following content is included on the **SECOND** page of your Assignment 3 submission.:

- What feedback did your tutor provide for Assignment 1?
- Briefly describe how you incorporated this feedback to improve your Assignment 3 submission (maximum 150 words).

3. **Assignment 3: Presentation (Slides + Verbal) + Peer-review Evaluation**

Weight: 10% of the unit mark

Submission format: one PDF file (Slides)

Size: a maximum of 10 slides (Slides)

You need to submit your presentation slides along with your final report. The 4 minute presentation is given in Week 12 during your assigned applied class and after your presentation, the tutor will ask at least one question to the presenter (1 minute). You will also be required to review and provide feedback on presentations of other students (peer-review) during the applied class in Week 12, using the Google Form provided.

How you will be assessed

See the marking rubric to understand how we will grade your assignments.

To introduce you to various important and novel project ideas developed by your peers and ensure a more accurate and fair assessment of your assignments, we will conduct peer grading for different parts of the assignments, as outlined below.

Assignment 1 proposal: The 15% awarded for your proposal is broken down into the following categories:

- **Problem Clarity** (2%): Is the problem well-articulated and clearly defined?
- **Business Model Analysis** (2%): Is the role of data in the project clearly articulated in relation to the business model? Are the benefits and value of the project clearly outlined? Are the primary stakeholders identified and their needs addressed?
- **Problem Importance** (4%): Does the project have real-world applications? Does it address key social, environmental, or business challenges and demonstrate potential for significant social impact?
- **Novelty** (4%): Does the project address an important and novel problem? Does it introduce a new or unconventional approach? Does it tackle an underexplored or emerging issue in data science?
- **Peer grading** (3%): You will review 6 randomly selected Assignment 1 submissions from other students and rate them based on **Problem Importance** and **Novelty**. Your peer-grading mark (3%) will be awarded in proportion to the number of reviews completed. Completing all 6 reviews will earn the full 3%. The peer-graded scores for **Problem Importance** and **Novelty** will be averaged and combined with the tutor's evaluation score to determine the final score for these aspects of a project. The average peer-graded score and the tutor-assigned score will each contribute equally to the final score. **Important:** *Please ensure that the final project text saved/submitted in FLoRA matches the PDF version submitted on Moodle, as the FLoRA submission will be used for peer grading, as detailed later. You may save/submit the written report multiple times, and the final version saved/submitted will be exported for peer grading.*

Please ensure that

Assignment 3 report: You will be assessed on your ability to:

- define the problem, provide background and significance, outline specific goals, analyze the business domain and its value creation, identify key stakeholders and their benefits, summarize existing research or industry solutions, highlight gaps in current approaches, and justify the project's novelty and potential impact (You can reuse the content from Assignment 1 for this section);
- discuss potential data sources and analyze their characteristics (e.g., the 4 V's) and evaluate the required platforms, software, and tools for data processing and storage based on the specific characteristics of the data or consider potential options (e.g., platforms, software, and tools) if your project expands in the future;
- specify the data analysis techniques and statistical methods (e.g., decision tree or regression tree) applicable to the project. Provide a rationale for the selected methods and discuss the expected high-level outcomes;
- identify a usable dataset for the proposed project and perform some basic analysis on the identified dataset to demonstrate the feasibility of the project, using R (e.g., detailing the information/features contained in the dataset, analyse the basic characteristics of the dataset, etc.), and report the analysis process and result in the demonstration section of a final report;
- describe any standards used in your data science process and practices for data governance and management in the project, e.g., how to address key issues such as data accessibility, security, and confidentiality, as well as potential ethical concerns related to data usage;
- think critically and creatively, providing justification and analysis;
- provide a good quality of report in terms of structure, expression, grammar and spelling.

For both assignments, make sure that any resources you use are acknowledged in your report. You may

need to review the [FIT citation style](#) to make yourself familiar with appropriate citing and referencing for this assessment. Also, review the [demystifying citing and referencing](#) guide for help.

Please also make sure that the Turnitin scores will be generated properly for your submissions. If a submission receives a high Turnitin score (e.g., more than 15%), the student will likely need to provide further explanation on the project idea and a penalty might be imposed on the submission in case no proper justification is provided.

If you use GenAI for this assignment (except for discussing potential project ideas with the GenAI-powered chatbot in FLoRA or seeking feedback from the chatbot), you must clearly document the type of GenAI used, how it contributed to the assessment, and provide a written acknowledgment of its use and extent in the final report.

Assignment 3 Presentation (Slides + Verbal Presentation + Peer-Review Evaluation): The 10% awarded is broken down into the following categories:

- Presentation (Slides) – 2% (evaluated by your tutor);
- Presentation (Verbal Presentation) – 3% (evaluated by your tutor);
- Peer-Review Evaluation – 5% (average scores given by your peers in the same applied class during Week 12). You may only evaluate projects from other students in your class and are not allowed to evaluate your own project.

What you need to do

Before you begin, make sure you:

- You are highly recommended to review the “inspiring” materials provided in FLoRA to select a topic that you would like to work on. Also, you are highly recommended to propose your own interesting and novel topic and please feel free to discuss it with your tutors to ensure the topic is suitable.
- Download the **marking rubric** (available on Moodle) as guidance on how you will be assessed.

Choose a data science project topic, and then:

1. **Do preliminary research** about your project topic and the relevant technologies by conversing with the GenAI-powered chatbot
2. **Write and submit your proposal** with cited references (Assignment 1)
3. **Research and prepare your final report** with cited references.
4. **Submit your report and do a presentation** (Assignment 3).

You are free to modify the initial proposal sections submitted for Assignment 1 (especially in response to feedback from your marker), or even change topics, when you are working on Assignment 3.

How to Submit

Once you have completed your work, take the following steps to submit your work. Penalties may be applied to your marks if the following instructions are not followed.

1. For Assignment 1, please finish and save the project proposal first in FLoRA, copy & paste it into a word processing tool (e.g., Microsoft Word) for the purposes of structuring/formatting, then save the project proposal in the PDF format and submit it on Moodle.
2. Please ensure you **name the file containing your proposal/report/slides** correctly using the following format:

FirstName_StudentNumber_AssignmentNumber(_report or _slides).pdf
e.g., Guanliang_12345678_Assignment1.pdf or

Guanliang_12345678_Assignment3_report.pdf or

Guanliang_12345678_Assignment3_slides.pdf

3. Upload your assignment file in the corresponding assignment link provided on Moodle.

Those unable to attend week 12 applied class for presentation:

Those who cannot do the presentation in their original applied class and would like to attend another applied class to do the presentation, please first contact the tutors of the class that you would like to attend to check whether the tutors have additional capacity to accommodate you and if yes, then you can join the class for presentation. We cannot guarantee that you will be admitted to another applied class for presentation if you cannot attend your original class due to the limited capacity the teaching team has.

Those who cannot do the presentation in any applied classes due to valid mitigating circumstances, can record and submit a video 4 minutes duration, through Youtube along with slides. **Please notice that, if you choose to do a video presentation, that means you will lose the 3% verbal presentation as well as the 5% peer-review evaluation in Assignment 3, i.e., only a maximum of 2% will be given for Presentation (Slides + Verbal Presentation + Peer-review Evaluation) if you fail to give a verbal presentation.** We recommend you produce a video by doing a 4 minute screen capture of your slides with voice over entered concurrently via microphone. You must upload your video through Youtube (make it unlisted, not private), and provide a LINK to your Youtube video in your submission (on the first page of the presentation slides) on Moodle. However, DO NOT include your video on Moodle as part of the submission. Please check the details of this, and confirm with a lecturer and your tutor a week in advance of presentation week.

Further advice on the assignment:

Here is some further advice from the teaching team regarding the assignment:

1. Make sure to carefully read the assignment specification above.
2. The project should be data-centred -- ideally combining multiple sources of data to develop your own project that can solve a real-world problem.
3. The project should contain a clear statement of the problem being tackled. What is the objective/purpose of the project?
4. Ensure that the project's novelty and benefits are clearly communicated. What makes the project novel? Will it provide financial benefits or contribute to social good?
5. The report needs to be "telling a story", and to be convincing somebody to "invest in your project" so that it can be built.
6. Try not to make the project too broad. It should be an achievable data science project.
7. Read up as much as you can on the particular topic you've chosen in order to be able to describe the data (and software) requirements of the project.
8. Make it clear where the data would come from for the project:
 - o Is the data proprietary? How would it be collected?
 - o If the data is public, you should do some exploratory data analysis on it.
9. What preprocessing would be needed? How would the data need to be preprocessed before it can be used? What software might be needed? Can the preprocessing be distributed?
10. Finally, make sure you've seen the set of possible section headings suggested above and structure your report accordingly.