Assignment 1: Workbook

Use this template document to compile your responses to the weekly assignment tasks.

# Task 1

This task is based on the learning materials in Week 2.

Study design and data collection

**Step 1: Download** this document. You have already completed this step.

**Step 2: Find** a past or ongoing example of data collection. You will need to find a reference that gives sufficient information about the design of a data collection example to answer the prompts in Step 3. You may wish to visit the following resources:

* [Australian Bureau of Statistics surveys of people, families and households](https://www.abs.gov.au/participate-survey/household-survey)
* [Australian Bureau of Statistics survey of business and organisations](https://www.abs.gov.au/participate-survey/business-survey)
* [UC Urvine machine learning repository](https://archive.ics.uci.edu/)
* [Australian clinical trials](https://www.australianclinicaltrials.gov.au/anzctr_feed/form)

**Step 3: Analyse** the study and respond to the following prompts:

* Give a publicly accessible reference to the study. You may wish to use [QUT cite|write: APA](https://www.citewrite.qut.edu.au/cite/qutcite.jsp#apa) for support with referencing.
* Briefly describe the purpose of the study.
* Name the type of study using at least one term from this week (e.g. observational study, survey, clinical trial, experiment, longitudinal study, etc).
* Identify at least one positive aspect of the design using concepts from this week. Discuss how this aspect might positive impact the study.
* Identify at least two potential sources of bias. Discuss how these potential sources of bias might impact the study.

**Step 4: Write** your responses as a reflection in this Assignment template - in 150-300 words. See the activity page for an example of a written reflective response.

**Step 5: Save** this document ready to add your remaining task responses in the following weeks.

# Task 2

This task is based on the learning materials in Week 3.

Ethical issues related to handling and processing data

**Step 1: Open** your recently updated Assignment template. You have already completed this step.

**Step 2: Read** the scenario listed on the activity page.

**Step 3: Consider** the following prompt question to help Jennifer in her ethical decision-making.

* Do you think Jennifer should present the analysis based on the merged data as her supervisor requests or not? Why or why not?

**Step 4: Evaluate** both options from two of the ethical lenses from the following list:

* Rights Lens
* Justice Lens
* Utilitarian Lens
* Common Good Lens
* Virtue Lens
* Care Ethics Lens

**Step 5: Write** your responses to Step 3 and 4 in 200-400 words.

**Note**: For the sake of brevity, do not expand on other possible actions in response to the problem. Simply assess the action of presenting the analysis based on the merged data or not through two ethical lenses.

**Step 6: Save** your written response in this document ready to add your remaining task responses in the following weeks.

# Task 3

This task is based on the learning materials in Weeks 4 and 5. It is divided into 4 parts.

Sources of error and bias in data analysis and interpretation

Task 3.1 Scenario - The Plummeting Profits of QUT Mart

This task is based on the learning materials in Weeks 4 and 5. This is part 1 of 2.

**Step 1: Open** your recently updated Assignment template. You have already completed this step.

**Step 2: Read** the scenario about QUT Mart and their experiences implementing automation of an inventory management system on the activity page.

**Step 3: Consider** the scenario and respond to the following prompts.

* Do you think the analysts have made any interpretation errors related to the following concepts:
* McNamara’s Fallacy
* Conditional Probability
* Correlation vs Causation
* Regression to the Mean
* Questionable Research Practices
* If any, describe how these interpretation errors may have occurred.

**Step 4: Write** your responses to the prompts in 100-200 words.

**Step 5: Save** your written response in your assignment template document ready to add your remaining task responses on the following activity page.

## Task 3.2 Analysis in Society – Interpretation Errors in Your World

This task is based on the learning materials in Weeks 4 and 5. This is part 2 of 2.

**Step 1: Open** your recently updated Assignment template. You have already completed this step.

**Step 2: Consider**the interpretation errors or pitfalls that have been discussed in this week's learning materials and select one:

* McNamara’s Fallacy
* Conditional Probability
* Correlation vs Causation
* Regression to the Mean
* Questionable Research Practices

**Step 3: Find** a real-world example of this error or pitfall occurring within a career related to your field of study.

**Step 4: Respond** to the following prompt:

* Briefly discuss the error in your real-world example, explaining how it occurred.

**Step 5: Write** your response to the prompt in 100-200 words.

**Step 6: Select**one ethical lens that you learned about in this unit (See Module 1).

**Step 7: Analyse**your real-world example in light of your selected ethical lens and respond to the following prompts:

* Describe how any conclusions or decisions based upon the data analysis in your example may be viewed through this lens.
* Identify, if any, potential societal implications of these decisions or conclusions.

**Step 8: Write** your responses to the prompts in 100-200 words using the Pros and Cons form below.

**Step 9: Save** your written response in your assignment template document ready to add your remaining task responses in the following weeks.

Pros

Text here.

Cons

Text here.

# Task 4

This task is based on the learning materials in Week 6.

## The Data Science Pipeline in action

**Step 1: Open** your recently updated Assignment template. You have already completed this step.

**Step 2: Select** a scenario that either applies to your field of study or everyday life. Your scenario should be one of the following two options:

1. A scenario/service/company within your field of study.
2. One of the following suggested options:

* a social media service
* a grocery store
* a general practitioner doctor’s office
* a clinical trial

**Step 3: Consider** the 'Data Science Pipeline' and how this might apply to your chosen scenario.

**Step 4: Respond** to the following prompts:

* Briefly describe the scenario – does it provide a service? Is this a public-facing business? etc.
* The first component of the data science pipeline is data input & design. For your chosen scenario, what is the source of data? Where does the data come from? If it comes from sampling, what about the sampling design needs to be considered?
* Moving onto handling & processing, which data is being collected? Are there data that are not retained, if so why not? Are there any pre-preprocessing steps that would be useful to apply to the data?
* For analysis & interpretation of these data, what would anyone wanting to use this data have to keep in mind? Note: This data could be used internally by the collector, externally by others, or both. Think of all applicable scenarios.
* Could this data be open? Should FAIR principles apply? What about CARE? Could you sell these data? If sold, what sort of insights/value could be derived from this data? (See the activity page for the definitions of these key terms).

**Step 5: Write** your responses to the prompts.

**Step 6: Save** your written response in your assignment template document ready to submit your assignment before the due date.