

Assignment 1: Part A (Question Formation and Exploratory Analysis)

Start Assignment

- Due 15 Jun by 23:59
- Points 25
- Submitting a file upload
- Available until 15 Jun at 23:59

Important Details

Weighting: 25%

Due: End of Week 3, Sunday 11.59 pm

Overview

Please refer to the [Project Overview \(https://myuni.adelaide.edu.au/courses/105232/pages/project-overview\)](https://myuni.adelaide.edu.au/courses/105232/pages/project-overview)

Task Description

Milestone 1A: Task Overview

- Formulate question(s) and identify, inspect, collate, clean and consolidate data that will help answer those questions. In finding and inspecting data you are expected to assess its adequacy for answering the questions. You should refine your questions or seek more data in response to these findings.

Milestone 1A: Task Description

- Prepare a document that:
 - Outlines and describes **initial questions** of industrial or broader societal relevance that can be answered with the data.
 - Defines the sets of data that can be collated, cleaned, combined and used to answer the question. This data must be **Big Data** in the sense described in [Module 1 \(https://myuni.adelaide.edu.au/courses/105232/pages/module-1-overview\)](https://myuni.adelaide.edu.au/courses/105232/pages/module-1-overview). This means the data might be large in **volume** or **complexity** and/or come from **multiple sources** or be in **multiple forms** (e.g., text, tables, images) or **multiple formats**.
 - Describes how you processed your data and your assessment of how it will help answer your question(s). Identify possible **deficiencies** and **pitfalls** in the data and describe how you addressed these or how these might lead you to **refine your questions**.

- Describes your **refined question(s)**, and outlines what you plan to do next to analyse the data and describe at least **one backup question** and data source that would allow you to progress the project.
- Represents an original question not answered elsewhere to the best of your knowledge.
- To do this, you will need to do the following tasks (i.e., separate to the report):
 - **Identify** your question(s) and **identify** your potential data sources.
 - You will need to **assess** the suitability of available big data to answer your question(s).
 - You will need to **identify** alternative data sources and alternative question(s).
 - You will need to **collate** the data so that it is amenable to further analysis and visualisation in later stages of the project.
 - You will also need to make a **preliminary plan** for the tools you will use in the next stage.

Note: your question(s) should be defined broadly enough so that they can be easily refined or modified if the required data is not available. You should **identify** back question(s) and **identify** alternative (perhaps related) data sources.

Note: as the course progresses you will be exposed to more skills for processing different types of data and the set of possible data sources will expand.

Referencing

- For this assessment you need to include references for the data and references relevant to the question.
- You should have at least four references.
- The referencing style for the Master of Data Science is **Harvard** referencing style.

Submission

You will submit this as a text document (doc, docx, pdf) via the link at the top of this page.

Examples:

These are the special guides to assist you:

- [Creating a Successful Project \(https://myuni.adelaide.edu.au/courses/105232/files/16625590?wrap=1\)](https://myuni.adelaide.edu.au/courses/105232/files/16625590?wrap=1)  (https://myuni.adelaide.edu.au/courses/105232/files/16625590/download?download_frd=1)
- [Project Exemplars \(https://myuni.adelaide.edu.au/courses/105232/files/16625593?wrap=1\)](https://myuni.adelaide.edu.au/courses/105232/files/16625593?wrap=1)  (https://myuni.adelaide.edu.au/courses/105232/files/16625593/download?download_frd=1)

Academic Integrity Policy

By submitting an assignment via MyUni/Canvas, you are agreeing to the following statement:

I declare that all material in this assessment is my own work, except where there is clear acknowledgement and reference to the work of others. I have read the University of Adelaide's [Academic Integrity Policy](http://www.adelaide.edu.au/policies/230) [_\(http://www.adelaide.edu.au/policies/230\)_](http://www.adelaide.edu.au/policies/230).

I give permission for any assessed assignments to be reproduced and submitted to other academic staff for the purposes of assessment and to be copied, submitted and retained in a form suitable for electronic checking of plagiarism.

[Applicable policy links](https://www.adelaide.edu.au/policies/700) [_\(https://www.adelaide.edu.au/policies/700\)](https://www.adelaide.edu.au/policies/700)

Questions?

Please post your questions on the [discussion board](https://myuni.adelaide.edu.au/courses/105232/discussion_topics) [_\(https://myuni.adelaide.edu.au/courses/105232/discussion_topics\)_](https://myuni.adelaide.edu.au/courses/105232/discussion_topics).

Assessment 1A: Big Data Analysis

Criteria	Ratings					Pts
Part 1: Problem description (Task 1)	4 Pts Full marks Clearly describes one or more initial questions of industrial or broader societal relevance that can be answered with data.	3 Pts Partial marks Describes one or more initial questions of industrial or societal relevance but has some elements that need more clarification and/or links to data.	2 Pts Partial marks (Pass) Questions are described for industrial problems that could be answered by data but the discussion is insufficient.	1 Pts Partial marks Questions described for industrial problems could be answered by data, but some important questions are missed.	0 Pts No marks Missed this part, or the questions are not for industrial applications, or the questions could not be solved by data.	4 pts
Part 2: Dataset description (Task 2)	4 Pts Full marks Precisely defines the sets of data that can be collated, cleaned, combined and used.	3 Pts Partial marks Defines the sets of data which are mostly feasible but some questions remain to be addressed.	2 Pts Partial marks (Pass) The datasets are aligned for a big data analysis task in the industry, but feasibility of using the data is not properly addressed.	1 Pts Partial marks The datasets are correct for an industry task, but they are not suitable for big data analysis.	0 Pts No marks The datasets are inappropriate or incorrect for big data analysis.	4 pts
Part 3: Initial Data processing (Task 3)	7 Pts Full marks Clearly describes how the data is processed and how it will help answer the questions. Clearly identifies possible deficiencies and pitfalls in the data and provides clear description on	5.25 Pts Partial marks Clearly describes the data processing steps and the possible deficiencies and pitfalls of the data. But some aspects of how to address them are missing.	3.5 Pts Partial marks (Pass) Describes the data processing steps and some possible deficiencies and pitfalls of the data, but misses or does not provide description on how to address them.	1.75 Pts Partial marks Describes the data processing steps but misses steps and/or misses the discussion of data issues and solutions.	0 Pts No marks Lacks of discussion of the data processing steps.	7 pts

Criteria	Ratings					Pts
	how to address them.					
Part 4: Refined problem and Plan (Task 4)	5 Pts Full marks Has clearly defined questions and provides a clear and detailed plan for the next step data analysis. Able to describe at least one backup question and data source that would allow the project to progress.	3.75 Pts Partial marks Clearly defines the questions and has a plan for next steps in the data analysis, but more detail could be provided. A backup question and data source is provided but needs further explanation.	2.5 Pts Partial marks (Pass) Refines the questions and has a plan for next steps of data analysis but not in sufficient detail. Provides backup question(s) and data source without explanation.	1.25 Pts Partial marks Refines the questions and has some basic discussions of the plans for next step data analysis, but not in sufficient detail. Does not provide backup question(s) and data source.	0 Pts No marks Backup question(s) and data sources are missing, or the questions are incorrect and the data source is not suitable.	5 pts
Part 5: Writing and references	5 Pts Full marks The report is written in English with good structure and logic with minimal typos or grammatical issues. Provides a comprehensive list of reference sources using Harvard referencing style.	3.75 Pts Partial marks The report is written with good structure and logic, but with some issues in written English or referencing format issues.	2.5 Pts Partial marks (Pass) The report is written with reasonable structure and logic, but has many issues with written English or misses important references. References may not be referenced correctly.	1.25 Pts Partial marks The report is written either with bad structure, logic or problematic written English. References may be absent or not correctly referenced.	0 Pts No marks Missing submission. Submission is not written in English or the writing is hard to understand.	5 pts
Total points: 25						