**Assessment Task 2: Data analysis project**

**Weight: 50% (see below for breakdown to different parts)**

**Rationale**

As a data scientist one of your core duties will be to work with a team of people to analyse complex datasets, and to report back the results of that analysis to stakeholders from a variety of backgrounds, who often have different needs and capabilities. This assessment task will give you a chance to experience the complexity that can often arise in this situation.

**You must keep a timestamped reflective journal throughout your groupwork project.**

You **MUST** do this using the private reflective journal discussion forum that has been set up for you in Canvas (you can find it under the Discussions tab). NB. Marks will be lost if your journal is not recorded throughout the group project!

**Part A (Proposal)**

**Weight:** 10%

**Length: 750-1000 words (not including code samples in appendix)**

**Group Assessment**

**Submitted: On Canvas in PDF format by *one group member*. Be sure to follow the naming convention defined at the front of this brief.**

**Task**

In this Assessment item, you will work in a team to produce a project proposal for your data analysis project. You will do this by following this sequence of steps:

1. Define a broad research area of interest (e.g. public health, climate change, demographic change, finance etc.) and form groups accordingly.
2. Work to define a set of well specified research questions for your broad area of interest. (Note: at least some of these should be actionable – see Part B).
3. Look for a range of datasets that might help to answer these questions.
4. Refine at least one research question/s so that it can be answered by a **Regression Model**. (Note: you can ask other questions too, but you must produce at least one regression model in Assessment 2B.)
5. Write a proposal that summarises the following:
   1. the rationale and stakeholders for the project,
   2. the research questions,
   3. the range of datasets examined as well as those chosen for the analysis (include details about how you merged the different datasets and an assessment on whether the granularity of the data sources is sufficient to answer your research questions),
   4. the regression modelling techniques to be employed and,
   5. any issues that you anticipate might arise in carrying out the project.
   6. Include an Appendix that contains code samples demonstrating the data acquisition and merger processes that you have used to date.

**Assessment Criteria: Part A**

|  |  |  |  |
| --- | --- | --- | --- |
| **SLO** | **CILO** | **Assessment Criteria** | **Weight** |
| 2 | 1.4 | Clarity in articulating the research questions along with a well defined proposal for making the invisible visible for a specified set of stakeholders. | 30% |
| 4 | 1.2 | Level of expertise using key R functionality demonstrated in the process of data acquisition, and creativity in solving the problem of finding and merging datasets that can answer the research questions. | 50% |
| 3 | 4.2 | Eloquence and robustness of the argument used to justify the proposal. | 20% |
| **Sub Total** | | | **100** |
| **Total (10%)** | | | **/10** |

**Part B (Presentation and Report)**

**Weight:** 20%

**Length:**

* **Presentation: 6mins**
* **Report: 5000 words (not including code which can be included in an appendix)**

**Group Assessment**

**Task**

Groups will now work in teams to implement the project proposed in Part A (taking on board feedback received from the teaching team).

Different stakeholders often have different expectations as to how statistical information and models will be communicated to them. Thus, senior managers (who are frequently the final decision makers) often expect a brief presentation, but will rely upon a separate set of recommendations from an in house team of people who are more expert in a domain. These recommendations are often derived from a combination of reports and presentations. Your challenge in this part of the project will be to communicate with both sets of stakeholders:

1. Presentations will be made during an on campus session. They should be pitched for an audience of senior decision makers who are not necessarily expert in the domain (e.g. a CEO or CIO).
2. You will have a short time, and one slide, to define your target audience before your presentation begins, and identify the problem that your analysis seeks to address.
3. Feedback will be given from your cohort and members of the teaching team.
4. Reports should then attempt to incorporate ideas provided through feedback received from the rest of the class.

**Assessment Criteria: Part B (Presentation)**

|  |  |  |  |
| --- | --- | --- | --- |
| **SLO** | **CILO** | **Assessment Criteria** | **Weight** |
| 3 | 4.2 | Quality and professionalism of presentation, and responses to questions, that are well tailored for your specified non-technical audience. | 30% |
| 2 | 1.4 | Strength of rationale and framing for research questions and preliminary results (including graphical and numerical summaries), along with clearly delineated limitations of the proposed study. | 30% |
| 6 | 4.3 | Effectiveness in presenting the exploratory data analysis to the defined stakeholder that enables rapid sensemaking and prompts new insights. | 40% |
| **Sub Total** | | | **100** |
| **Total (10%)** | | | **/10** |

**Assessment Criteria: Part B (Report)**

|  |  |  |  |
| --- | --- | --- | --- |
| **SLO** | **CILO** | **Assessment Criteria** | **Weight** |
| 5 | 2.2 | Appropriateness of statistical models with assumptions and mitigation strategies clearly justified. | 30% |
| 1 | 2.4 | Clarity and strength of alignment between the exploratory analysis, models generated, features chosen, and their limitations. | 40% |
| 3 | 3.1 | Clarity and fluency in communicating your findings to a technical target audience. Soundness of the model interpretation and implications, and professionalism of the executive summary for C-suite decision makers. | 30% |
| **Sub Total** | | | **100** |
| **Total (10%)** | | | **/10** |