ASSESSMENT BRIEF

36103 Statistical Thinking for Data Science

Assessment

This document outlines each assessment task for STDS, including key dates, submission formats, weightings, assessment briefs, and associated Subject Learning Objectives (SLO) and Course Intended Learning Outcomes (CILO). Students should familiarise themselves with this document and refer to it throughout each assessment task. For help, tips, and more information, head to Canvas and look for each assessment task in the 'Modules' tab.

1. 36103 Assessment Overview

Assessment summary table with key dates:

Deliverable	Description	Type	Weight	Due
Assessment Task 1: Exploratory data analysis	Exploratory data analysis report	Individual	20%	11.59pm Sunday 8th September 2024
Assessment Task 2: Data analysis project	Team expectations agreement	Group	0%	11:59 pm Sunday 15th September
	Presentation	Group	15%	Saturday 12 October online
	Report	Group	15%	11:59 pm Sunday 20 October 2024
Assessment Task 3: Data analysis project for marketing campaigns	Report	Individual	50%	11.59 pm Sunday 10 November 2024

Assessments are a blend of individual and team-based work.

2. Additional information

Citations: Proper referencing is mandatory (APA style preferred) for all externally sourced material.

Submission Requirements:

- All assignments need to be submitted via Canvas unless otherwise instructed
- Submissions must include a title page with the subject, assessment task, student name(s) and IDs, date of submission, and the title of the assessment
- Please use the following file naming format for each submission:
  - For individual assignments: StudentName_AssignmentName_Date
  - For group assignments: TeamName_AssignmentName_Date
- Please use embedded objects instead of linked objects for content sourced externally.

Length penalties:

Submissions exceeding task Length by more than 20% will be penalized (10% of the overall assessment mark). Tables, figures, references, and appendices are not included in word limits.

Plagiarism:

Assessments will be checked against both other students' work and external sources for evidence of plagiarism. Any confirmed cases of plagiarism may result in reduced marks for an assessment.

Extensions and late penalties:

If unavoidable circumstances arise, students can apply ***before*** the assignment due date for an extension of up to five days by sending an email to the subject coordinator. This needs to outline the reason they are unable to submit on time and include an outline of how far they have progressed with the assignment.

- Extension requests submitted after the due date will only be considered in exceptional circumstances; note that work, travel or hardware issues are not valid reasons.
- Unless formal extension dates are agreed upon in writing each late submission will be penalised 10% per day after the due date (to a Pass grade).

Special Consideration:

Extensions of more than five days require a formal application for Special Consideration in accordance with university policy.

Minimum requirements:

To meet the minimum requirement for the course, students must attain a minimum of 50% marks to pass.

Assessment task 1: Exploration of data skills and issues

Task: This assessment is intended to conduct exploratory data analysis (EDA) on a marketing campaign dataset from a telecommunication company. A telecommunication company recently launched a marketing campaign to promote the adoption of their new subscription plan among customers. The company seeks assistance in gaining a comprehensive understanding of their customers and identifying the customer segments that display the highest responsiveness to marketing campaigns. The dataset is available on Canvas. The response variable, subscribed, indicates whether the client subscribed to a new plan, which was the objective of the campaign.

The dataset may have issues such missing information and data errors. Identifying and handling such issues is part of the assessment.

The requirements involve applying a minimum of three distinct exploratory data analysis techniques to gain preliminary insights from the data.

➢ For the assessment task, you are required to submit a report that includes the following elements as a minimum requirement:

➢ Problem formulation. It should incorporate a comprehensive discussion of the analysis context, the specific problem at hand, the pertinent questions, and hypotheses to be addressed. (Criteria 1: CILO 2.2)

➢ Data preprocessing. Your report should encompass a comprehensive and informative overview of the dataset. It is essential to ensure that the dataset is error-free and undergoes accurate processing before analysis. Furthermore, it is crucial to provide a concise description of the data processing steps undertaken. (Criteria 2: CILO 2.4)

➢ Exploratory data analysis (EDA). It involves a comprehensive examination of relevant variables both individually and in relation to each other. This is achieved through the use of appropriate figures and descriptive statistics. During this process, you take note of all data characteristics that are pertinent for model development. The results of the EDA are clearly explained in terms of their relevance to the overall goal of the project. (Criteria 2: CILO 2.4)

Finalise and submit the report:

o Before submission, check for spelling and grammatical errors. (Criteria 3: CILO 3.1)

o Format the report to enhance readability, use headings and choose appropriate fonts, etc. (Criteria 3: CILO 3.1)

o Submit the report in Canvas in pdf format.

Assessment Criteria:

	Criteria	Weight (%)	SLO	CILO
1	Clarity and brevity in explaining data issues, and the appropriateness of exploratory data analysis.	40	Justify statistical choices	Exploring, interpreting and visualising data
2	Depth of insight and applying a minimum of three distinct exploratory data analysis techniques to gain preliminary insights from the data.	50	Apply statistical methods	Designing and managing data investigations
3	Clarity and fluency in communicating your findings to a technical target audience.	10	Justify statistical choices	Developing strategies for innovation

Length: A maximum of 7 pages

Due: 11.59 pm Sunday 8th September 2024

Assessment task 2: Data analysis project

This task focuses on exposing students to each key step in a data science project cycle using real-world data. Students will work in teams to propose, execute, and critically reflect on a full data science research project. They will learn how to develop and pitch a research idea, execute their chosen project and data, and discuss ethical issues related to privacy, consent, and data sovereignty, especially when working with data related to Indigenous Australians. Additionally, students will critically reflect on the implications of this learning for their future professional practice, communicate findings both verbally and in report format, and self-evaluate project progress and areas for improvement. This task is divided into three subtasks.

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, and gain experience executing a full data science project cycle.

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. The aim of the assessment task is to communicate with both types of stakeholders.

Task 2:   Students work in teams of 5-7 people with complementary skills and backgrounds. Each team selects a context and works to define research questions that help them to propose, execute, and disseminate a data science project.

1. Team expectations agreement

This subtask does not count toward the final mark.
In the early stages of team formation, trust is often lacking, shared goals are absent, and roles remain unclear. To address these challenges, a team expectations agreement comes into play. This document establishes the team's purpose, expected outcomes, and outlines how the team will collaborate to achieve results. It serves as a set of agreements crafted during the initial formation of the team, ensuring alignment and shared understanding among all members from the outset.

The team expectations agreement should encompass the following elements:

   a. Group objective. We collectively agree to …
   b. Collaboration approach: We will define how we prefer to work together by…

2. Project presentation (group) worth 15%

   a. Students work in teams to carry out their proposed project.
   b. Projects are presented to the class.

   ➢ Prepare your group presentation (CILO 4.2, 1.4 + 4.3)

a. Presentations should briefly introduce the problem and relevance, methods employed, results, and conclusions.
b. Remember your target audience – non-technical decision makers.
c. Work together to develop a strong slide deck.
d. Think carefully about any data visualisations you include. There is a difference between technical visualisations to help you understand things vs visualisations to convey a message to an audience.
e. Practice your presentation as a group to ensure you stay on time. Practice runs will help things flow much better when presenting live.

➢ Present your results (CILO 4.2, 1.4 + 4.3)

Presentations will be made during an online session. They should be pitched for an audience of senior decision makers who are not necessarily expert in the domain.

➢ Assessment Criteria

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 data analysis to the defined stakeholder that enables rapid sensemaking and prompts new insights.	40%

3. **Project report (group) worth 15%**

Teams will write up their executed research project as a technical report. This report should include sufficient detail for the target audience to understand the project aim, assess the suitability of chosen datasets and methods, and critically evaluate the conclusions drawn. This includes model assumptions and evidence that they have been tested. Appropriate use of data visualisation and summary tables should be included to communicate key outcomes. Any externally source data or information should be correctly cited. An appendix containing the code used to undertake the research and produce outcomes should be included, and should be self-contained and well formatted.

a. Students work in teams to carry out their proposed project.
b. Project reports are submitted in written format.

➢ Work together to write up the technical report

o Ensure the report follows a logical structure. A useful format to follow would be; an executive summary, introduction, methods, results, discussion, and conclusion. (CILO 3.1)
o Define your research topic, aims, and objectives. (CILO 2.4 + 3.1)
o Describe and summarise the datasets used. (CILO 2.4)
o Use correct terminology and report appropriate statistics throughout. (CILO 3.1)
o Ensure you include at least one regression model. (CILO 2.2)
o Include visualizations and tables. Include figure/table legends and refer to them in the main text. (CILO 3.1)

- Include an executive summary at the start of the report, this should be a self-contained summary of the project and outcomes, aimed at non-technical readers, this forms a key part of CILO 3.1. (CILO 3.1)
- Be sure to cite any externally sourced information, including datasets. (CILO 3.1)
- Include a code appendix, following previous formats where code is self-contained and well formatted. (CILO 2.4 + 3.1)

➢ Finalise and submit the report

- Before submission, check for spelling and grammatical errors. (CILO 3.1)
- Format the report to enhance readability, use headings and choose appropriate fonts, etc. (CILO 3.1)
- Submit the report in Canvas in pdf format.

➢ Assessment Criteria

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 analyses, 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%

Length: Team expectations agreement: 1-2 Pages document

Presentation: 10-15 minutes Group presentation; Report: 500-700 words

Due: Team expectations agreement: 11:59 pm Sunday 15th September 2024

Presentation: Saturday 12 October online

Report: 11:59 pm Sunday 20 October

Assessment task 3: Data analysis project for marketing campaigns

Task: Assessment 3 expands on Assessment 1, making use of the same data set that is accessible on Canvas.

The objective of this assessment is to build data science models that yield valuable insights into the two crucial business questions: identifying the customer segments most responsive to marketing campaigns and deriving effective business strategies. To achieve this, you are required to:

- Develop statistical learning models that can predict the success of a marketing campaign for a given customer.

- Extract at least three insights that can assist the company in making informed decisions regarding their marketing campaigns.

Your report must show results for at least two different sets of predictions.

• At least one of your models should be a parametric model.

• At least one of your models should be a non-parametric model.

You should use at least one estimation method introduced in Module 3.

Your report should include the following elements:

- Problem formulation. It should incorporate a comprehensive discussion of the analysis context, the specific problem at hand, the pertinent questions, and the hypotheses to be addressed. (Criteria 1 and 2: CILO 1.4+2.2)
- Justification for model selection, including an explanation of the configuration and training choices made. (Criteria 2: CILO 2.2)
- A comparative analysis of the models, employing cross-validation or validation metrics. (Criteria 2: CILO 2.2)
- Parametric estimates and their corresponding interpretations. (Criteria 3 and 4: CILO 4.3+ 3.1)
- Proficiency in Data mining, demonstrated by the ability to extract relevant business insights from the data and effectively articulate them. (Criteria 3 and 4: CILO 4.3+ 3.1)

➢ Assessment Criteria

	Criteria	Weight (%)	SLO	CILO
1	Clarity in articulating the questions along with a well defined proposal for making the invisible visible for a specified set of stakeholders.	20	Formulate questions	Making the invisible visible
2	Soundness of the statistical methodology that also shows evidence of having applied relevant analytical methods to address the business questions.	35	Justify statistical choices	Exploring, interpreting and visualising data
3	Appropriateness of the interpretation applied to the results, where the resulting conclusions are well justified and answer the buisness questions.	35	Interpret results	Informing decision making
4	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 decision makers.	10	Justify statistical choices	Developing strategies for innovation

Finalise and submit the report:

- Before submission, check for spelling and grammatical errors. (Criteria 4: CILO 3.1)
- Format the report to enhance readability, use headings and choose appropriate fonts, etc. (Criteria 4: CILO 3.1)
- Submit the report in Canvas in pdf format.

Length: 700-1000-word Canvas Submission.

Due: 11.59pm Sunday 10 of November

Academic integrity

Where individual work is required for the purposes of assessment, the copying, unacknowledged use of, or reliance on the work of other individuals without acknowledgment is considered to be cheating/misconduct. The penalties imposed for cheating/misconduct or allowing work to be plagiarised are severe under the University rules and regulations.

Statement on copyright

Please remember that teaching materials and resources provided to you at UTS are protected by copyright. You are not permitted to re-use those for commercial purposes (including in kind benefit or gain) without permission of the copyright owner. Improper or illegal use of teaching materials may lead to prosecution for copyright infringement.

Statement on plagiarism

Plagiarism is one of the most serious crimes in the academic community. It indicates an attempt by someone to pass off the words and/or ideas of another as their own. To take any but a few sequential words of another without acknowledgment is plagiarism and tantamount to cheating.

Acts of plagiarism are penalised.

Statement on intellectual property

All courses in TD School involve industry, public sector, not-for-profit, and/or community engagement. In general, undergraduate students own their intellectual property (IP) rights as outlined in the UTS Student Rules and the UTS Intellectual Property Policy. Depending on the nature of the industry, public sector, not-for-profit and/or community engagement, students acknowledge that they may be required by these external partners to assign their IP rights in a project if they choose to participate in certain industry-related experiences. In situations where students have agreed to assign IP rights, UTS encourages our Industry partners to credit students by name if they elect to implement any student ideas (at a minimum).

In certain subjects, an organization may choose to share confidential company information for student learning purposes, e.g. data, insights, and reports. In such circumstances, students are required to return all physical and electronic copies of data and reports back to the organization and to share their insights, ideas, and reports with the partner via the subject coordinator at the end of the subject. This arrangement is outlined in a non-disclosure agreement from the organization signed by students on the subject to which this situation applies.