COMP20008-S2 2025 Assignment 2

Key Elements of the Report

Methods

- Clarity of Research Question(s)
- Comprehensive understanding of the dataset, including its characteristics, limitations, included features, and missing/underrepresented data.
- Thorough exploration of alternative methods and <u>explicit justification of choices, with clear explanation of the decision-making process</u> across the following:
  - Data Pre-processing: Select and justify techniques that meaningfully contribute to subsequent analysis.
  - Correlation Analysis: Explore options, justify selection, and apply methods that connect meaningfully to the research.
  - Supervised Methods & Evaluation: Select, justify, and apply ML models; explore and explain parameter choices; justify evaluation methods and metrics.
  - Clustering & Profiling (groups of 4 only): Explore alternatives, justify chosen methods, and ensure meaningful connection to the research.

Analysis of Data and Results

- Exploratory Data Analysis (EDA)
  - Use descriptive statistics, correlations, and initial data patterns.
  - Clearly connect preprocessing choices to later sections, explaining their impact on subsequent analyses.
- Interpretation
  - Provide in-depth, <u>meaningful analysis of results with well-justified interpretations</u>.
  - Extract novel, thought-provoking insights and, where appropriate, design complementary analyses for further investigation.
  - <u>Justify the significance of observed patterns</u>, linking findings strongly to the research question.
  - Discuss potential real-world implications of the findings.
- Limitations
  - Thoughtfully acknowledge the dataset and methodological limitations.
  - Provide insightful, well-justified suggestions for improving the analysis.

Report Presentation & Communication

- Visualisations & Results Clarity
  - Use multiple relevant charts, graphs, and tables.
  - Ensure proper labelling, captions, readability, and alignment of visuals with key insights.
  - Visualisations should effectively highlight trends and support arguments.
- Report Structure & Writing Quality
  - Include an executive summary and all required sections.
  - Follow formatting instructions: correct font, layout, indenting, and sectioning.
  - Maintain correct report length.
  - Ensure smooth transitions, logical flow, and cohesion across sections.
  - Use visual aids effectively to complement the narrative.
  - Provide appropriate references in APA style

COMP20008-S2 2025 Assignment 2

Report Rubric

	Methods (6 Marks)	Analysis (10 Marks)	Report Quality (4 Marks)
H1 (80%+)	Research question(s) are clear and precise. Shows a comprehensive understanding of dataset characteristics, limitations, and missing/underrepresented features. Considers multiple methods in each section, with well-reasoned justification for chosen approaches. Machine learning section includes detailed exploration of model parameters, explanation of their influence, and rationale for final settings. Methods are applied correctly, clearly connected to research aims, and integrated across sections. Demonstrates insightful reflection throughout.	EDA is thorough and insightful, using descriptive statistics and correlations to reveal clear patterns and explicitly linking preprocessing decisions to later analyses. Interpretation shows deep, meaningful analysis, with novel insights, well-justified significance of patterns, strong connection to the research question, and consideration of real-world implications. Where appropriate, develops complementary analyses to extend findings. Limitations are thoughtfully and critically addressed, with insightful suggestions for improvement.	Visualisations are highly effective, with multiple relevant charts/tables, excellent labelling, captions, and readability. They highlight trends and strongly support arguments. Report is polished: executive summary and all sections included, formatting followed precisely, correct length maintained. Writing shows excellent flow, cohesion, and professional clarity, with visual aids seamlessly integrated into the narrative.
H2 (70%-79%)	Research question(s) are clear and relevant. Shows a strong understanding of dataset characteristics and limitations, though reflections may be less detailed. Explores some alternatives and provides generally sound reasoning for chosen methods. Machine learning section includes discussion of parameters with mostly clear rationale. Methods are mostly correct and linked to the research aims, with only minor gaps in integration or depth.	EDA is clear and systematic, using appropriate statistics and visualisations, with some explanation of preprocessing impact. Interpretation provides solid analysis and justified insights, with occasional originality and clear links to the research question. May mention real-world implications, but not in depth. Limitations are acknowledged with reasonable suggestions for improvement, though less critical or detailed.	Visualisations are clear and relevant, with proper labels and readability; they support arguments well but may lack some depth or polish. Report includes all major sections and follows formatting with only minor lapses. Writing is coherent and flows logically, with visual aids effectively used.
H3 (65%-69%)	Research question(s) are present but may lack clarity or depth. Shows a basic understanding of dataset with some overlooked aspects. Considers only limited alternatives, often presenting a single method. Justification is brief or generic. Machine learning section includes basic mention of parameters, but explanation is minimal or inconsistent. Methods are generally correct but may include minor errors or weak alignment to aims.	EDA is basic but adequate, showing descriptive statistics and patterns but with limited depth or weak links to preprocessing. Interpretation is general and somewhat superficial, with few novel insights and limited justification of significance. Links to research question are present but not well developed. Limitations are mentioned but briefly or generically.	Visualisations are adequate but may be limited in number, uneven in quality, or only partly aligned with insights. Labelling/captions may be inconsistent. Report covers most required sections and attempts formatting but with noticeable issues. Writing is understandable but may lack smooth transitions or cohesion. Visual aids are used but not fully integrated.

COMP20008-S2 2025 Assignment 2

	Methods (6 marks)	Analysis (10 Marks)	Report Quality (4 Marks)
P (50%-64%)	Research question(s) are unclear, vague, or poorly aligned. Shows a limited understanding of dataset and little reflection on limitations. Methodological choices are narrow or weakly justified. Machine learning section includes superficial or missing discussion of parameters. Applications contain errors or inconsistencies, with poor integration across sections. Minimal evidence of reflection.	EDA is minimal or incomplete, with gaps in descriptive statistics, correlations, or explanation of preprocessing impact. Interpretation is weak, mostly descriptive, with little or no justification of patterns, weak connection to research question, and no real-world implications considered. Limitations are acknowledged superficially or not meaningfully addressed.	Visualisations are minimal, unclear, or weakly linked to findings. Labelling and readability often missing. Report may have missing sections, poor adherence to formatting, or incorrect length. Writing is fragmented with weak flow, and visual aids are simplistic or underused.
F (less than 50%)	Research question(s) are absent or incoherent. Shows little or no understanding of dataset characteristics. Methods are inappropriate, unjustified, or missing. Machine learning section ignores parameter selection/exploration. Applications are incorrect or omitted, with no evidence of reflection or alignment. Major errors in most sections.	EDA is absent, incorrect, or incoherent. No meaningful analysis of patterns or connection to preprocessing. Interpretation is missing or incorrect, with no justification of results and no link to the research question. Limitations are ignored or wholly inadequate.	Visualisations are absent, irrelevant, or unreadable. Report is missing key sections, ignores formatting, and does not meet length requirements. Writing is incoherent, with no logical flow. Visual aids are absent or meaningless.