Report Rubric

Methods (5 Marks)

- Clarity of the Research Question(s)
- Demonstrated understanding of the dataset, including key variables, relationships, and potential challenges.
- Data Pre-processing:
  - A minimum of 1 appropriate and justified data pre-processing task
  - Effective application of the chosen technique(s), with explanations on how it improves the quality of the data for analysis
- Correlation Analysis:
  - A minimum of 1 robust and relevant correlation analysis task
  - Effective application and justification of the chosen technique(s)
- Supervised Methods and Evaluation:
  - A minimum of 2 well-justified machine learning models applied tailored to the research question(s).
  - Accurate and meaningful evaluation comparison of used machine learning techniques, with clear reasoning behind used evaluation techniques and metrics
- Clustering and risk profiling (for groups of 4 only)
  - A minimum of 1 robust and well justified clustering analysis task
  - Identification of meaningful and justified risk profiles with meaningful insights.

Quality of Results (3 Marks)

- Visualisations & Clarity of Results
  - Use of multiple relevant charts, graphs, and tables.
  - Proper labelling, captions, and readability of visualisations.
  - Visualisations effectively highlight key insights and trends.

Analysis of Data and Results (8 Marks)

- Exploratory Data Analysis (EDA)
  - Descriptive statistics, correlations, and initial insights into data patterns
- Interpretation:
  - Strong linkage between findings and the research question.
  - Justification of why observed patterns are significant
  - Extracting novel, thought-provoking insights from the data
  - Discussing potential real-world implications of findings
- Limitations:
  - Thoughtfully acknowledging and addressing limitations in data or methodologies
  - Insightful suggestion for analysis improvement

Report Quality (4 Marks)

- Executive Summary
- No missing section
- Overall writing quality
  - Exact use the instructed font, layout with proper use of indenting, and sectioning
  - Effective use of visual aids
  - Correct length of the report (10-12 pages)
- Logical Flow and Coherence
  - Smooth transition between ideas and sections
  - Cohesive flow between different sections

	Methods (5 marks)	Results (3 Marks)	Analysis (8 Marks)	Report Quality (4 Marks)
H1 (80%+)	Clear and well-defined research question. Strong understanding of the dataset. Data preprocessing, correlation, and supervised learning models applied effectively with clear justification. Clustering (if applicable) is well-executed and provides meaningful risk profiles.	Multiple relevant and well-designed visualisations. Clear labelling, captions, and high readability. Visualisations effectively support findings and highlight insights.	Strong exploratory data analysis with well-explained patterns and relationships. Findings are deeply linked to the research question, with insightful, real-world implications. Thoughtful discussion of limitations and improvement suggestions.	Well-structured report with an excellent executive summary. Clear, professional writing with smooth transitions and logical progression. Follows formatting instructions, with effective visual aids and correct length.
H2 (70%-79%)	Research question is clear, with a good understanding of the dataset. Data preprocessing, correlation, and supervised learning models are applied correctly with minor gaps in justification. Clustering (if applicable) is mostly well-executed.	Good variety of relevant visualizations. Mostly clear labelling and readability. Visuals support findings but could be improved in clarity or effectiveness.	Good exploratory data analysis with mostly clear insights. Findings are linked to the research question but could be deeper. Limitations and improvements are discussed but lack depth.	Good report structure with a mostly clear executive summary. Writing is clear but could improve in transitions or logical flow. Mostly follows formatting guidelines with minor issues.
H3 (65%-69%)	Research question is somewhat clear. Basic understanding of the dataset. Some issues with data preprocessing, correlation, or ML model justification. Clustering (if applicable) is present but lacks depth. Use of some evaluation methods and metrics with limited justification and/or with some mistakes.	Some relevant visualisations but lacks variety or effectiveness. Labels and captions are present but could be clearer.	Basic exploratory data analysis with some relevant findings. Linkage to research question is present but weak. Some discussion on limitations and improvements.	Basic report structure with some missing transitions or logical progression. Executive summary is present but lacks clarity. Formatting guidelines are followed but inconsistently.
P (50%-64%)	Research question is vague or weakly linked to the dataset. Data preprocessing, correlation, or ML models are applied but with errors or weak justification. Clustering (if applicable) is poorly executed.	Few visualisations, and some are unclear or not well-justified. Poor labelling or captions. Limited support for findings.	Weak exploratory data analysis with minimal insights. Findings are not well connected to the research question. Little discussion on limitations or improvements.	Poorly structured report with weak logical flow. Executive summary is unclear or missing. Formatting has multiple errors.
F (less than 50%)	Research question is missing or unclear. Poor understanding of the dataset. Data preprocessing, correlation, or ML models are incorrect, unjustified, or missing. Clustering (if applicable) is absent or not meaningful.	Minimal or no visualisations, or they are irrelevant, unclear, or misleading. No labels or captions. No meaningful support for findings.	No meaningful exploratory data analysis. Findings are unclear, unrelated to the research question, or missing. No discussion of limitations or improvements.	Report lacks structure and logical flow. Executive summary is missing or unclear. Formatting is not followed.