GA1 Marking Rubric

		HD	D	C	P	F
EDA methodology (35%)						
Merged dataset (15%)	Correct structure for the output file as specified in the specification	It is a must. Note that the output file should be in the pre-defined structure the same as the sample output file. The correctness of the output file is measured by • Correctness of file names • Correctness of attributes • Correctness of values				
	csv file (15%)	Achieve at least 95% on overall accuracy	Achieve at least 90% on overall accuracy	Achieve at least 80% on overall accuracy	Achieve at least 70% on overall accuracy	Achieve less than 70% on overall accuracy
Methodology (20%)	Data processing techniques (15%)	The report has demonstrated a good solution in a proper way. The methodology consists of all required steps (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct information, etc) for the tasks, and producing the correct output.	The report has demonstrated a reasonable solution. One key step (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct information, etc) is missing or leads to the incorrect output.	The report has provided a fair solution. Two key steps (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct information, etc) are missing or lead to incorrect output.	The report has provided a poor solution. Three key steps (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct information, etc) are missing or lead to incorrect output.	The report has provided a bad solution. Only one or two steps are demonstrated, which clearly cannot achieve the correct output.
	Code efficiency (5%)	Code is highly efficient, optimised, and performance considerations are well addressed.	Code is efficient with minor optimization needed. Performance considerations are mostly addressed.	Code is functional but not optimised. Performance considerations are minimally addressed.	Code is functional but inefficient. Performance considerations are not well addressed.	Code is inefficient, not optimised, and performance considerations are ignored.
EDA report (50%)						
Documentation (45%)	Introduction and conclusion (5%)	The report has a well written introduction and conclusion that gives enough information to understand the main content of the report easily.	The report has an introduction and conclusion that cannot cover all main content.	The report has provided a fair solution. Two key steps (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct information, etc) are missing	The report has provided a poor solution. Three key steps (e.g., using correct functions for input and output, developing correct regular expression, retrieving correct	The report has provided a bad solution. Only one or two steps are demonstrated, which clearly cannot achieve the correct output.

				or lead to incorrect output.	information, etc) are missing or lead to incorrect output.	
	EDA design (10%)	EDA design is well organised with logical and appropriate steps for easy reproduction.	EDA design has key steps involved that are possible for reproduction.	EDA design has key steps with limited explanation and justification.	EDA design follows the general steps only, no explanation or justification.	EDA design is poor and missing key steps without explanation or justification.
	Findings (10%)	Valuable findings are well selected and presented with appropriate explanation and justification supported by results and visualizations from the EDA methodology. These findings are highly relevant to the key insights and ML questions.	Valuable findings are given with some explanation and justification supported by the results and visualizations from EDA methodology. Some findings are highly relevant to the key insights and ML questions.	Selected findings are given with limited explanation and justification supported by the results and visualizations from EDA methodology.	Findings are given without a careful selection. Limited explanations or justification are given with some results or visualizations from EDA methodology.	Findings are of low quality without explanation or justifications.
	Insights and ML questions (20%)	10 valuable insights and 5 ML questions are given with connection to findings with potential contributions to the research domain or business context.	10 insights and 5 ML questions are given with connection to findings. But not all of them have potential contributions to the research domain or business context.	At least 8 insights and 3 value ML questions are given with connection to findings. But not all of them have potential contributions to the research domain or business context.	At least 5 insights and 3 ML questions are given with connection to findings. But not all of them have potential contributions to the research domain or business context.	Less than 5 insights or 3 ML questions are given with connection to findings. But not all of them have potential contributions to the research domain or business context.
Report structure and presentation (5%)	Report structure and discussions	The report has proper sections and subsections (e.g. Introduction, …, Conclusion, Reference). The report is written in a scientific way with clear presentation and easy to follow. Reference is relevant, complete and correct.	The report has proper sections and subsections (e.g. Introduction, …, Conclusion, Reference). The report is written in a scientific way, but the presentation needs improvement. Reference is relevant, complete and correct.	The report has the key sections only. The content is not organised in a logical way. The explanation or justification is brief and short to provide valuable information. Reference is incomplete but relevant.	The report has the key sections with limited explanation or justification. The content is not organised in a logical way. Reference is incomplete or irrelevant.	The report has poor/no sectioning, and it is not well organised. The explanation and justification are missing or irrelevant. No reference is included.
Video presentation (10%)						
Presentation Clarity (5%)		The presentation is between 6-10 minutes. The presentation is	The presentation is clear and organised with minor areas for improvement.	The presentation is generally clear but has some organisation issues.	The presentation is somewhat clear but lacks organisation and smooth	The presentation is unclear, poorly organised, and hard to

	exceptionally clear, well-organised, and easy to follow.			flow.	follow.
Content Coverage (4%)	Covers all key aspects of the analysis comprehensively and in detail.	Covers key aspects of the analysis with minor gaps in detail.	Covers most key aspects of the analysis but lacks detail in some areas.	Covers basic aspects of the analysis but misses significant details.	Fails to cover key aspects of the analysis and lacks significant detail.
Delivery and Engagement (1%)	Highly engaging delivery with clear speech, good pace, and strong audience connection.	Engaging delivery with mostly clear speech and good pace, minor engagement issues.	Adequate delivery with understandable speech, moderate pace, and limited audience engagement.	Delivery is somewhat engaging but has issues with clarity, pace, and audience connection.	Delivery is not engaging, unclear speech, poor pace, and fails to connect with the audience.
Generative AI tools Declaration (5%)					
Generative AI tools Declaration (HURDLE)	**Note that every member of the team should sign and include this declaration in the assessment submission. Otherwise, will receive 'Fail' for the A1.**				