

## GA1 Marking Rubric

			HD	D	C	P	F
Task 1 (8/35)	Parsing Raw Files						
	Output (50%)	Correct structure for the output files as specified in the spec	It is a must. Note that the output files should be parsable and also be in the pre-defined structure (i.e., having correct tag name, structure, delimiters and datatypes). Unparsable outputs and the output files that have any deviations from the specification, will receive ZERO marks for the output.				
		Json file(40%)	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
		csv file(10%)	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 (35%)	Data processing techniques(20%)	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(15%)	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.
	Documentation (15%)	Report structure and discussions(10%)	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly. The report can be improved.	The report has some sectioning but it is not well organised. The explanation of the report is limited	The report has some sectioning but it is not well organised. The explanation of the report is poor.	The report has poor/no sectioning and it is not well organised. The explanation of the report is poor.
		Code structure and documentation (5%)	Code is well-structured, easy to read, and follows best practices. Code is thoroughly documented with clear and concise comments.	Code is generally readable but has structural issues. Deviations from best practices are noticeable. Code is well-documented with minor gaps in comments.	Code is somewhat readable but lacks structure. Deviations from best practices are noticeable. Code has minimal documentation and comments.	Code is somewhat readable but poorly structured. Deviates significantly from best practices. Code has few comments and lacks proper documentation.	Code is poorly structured and hard to read. Does not follow best practices. Code lacks documentation and comments

<b>Task 2 (12/35)</b>	<b>Text Preprocessing</b>						
	<b>Output (50%)</b>						
		Correct structure for the output files as specified in the spec	It is a must. Note that the output files should be parsable and also be in the pre-defined structure (i.e., having correct tag name, structure, delimiters and datatypes). Unparsable outputs and the output files that have any deviations from the specification, will receive ZERO marks for the output.				
		unigram (30%)	Achieve at least 90% on overall accuracy	Achieve at least 80% on overall accuracy	Achieve at least 70% on overall accuracy	Achieve at least 60% on overall accuracy	Achieve less than 60% on overall accuracy
		bigram (5%)	Achieve at least 90% on overall accuracy	Achieve at least 80% on overall accuracy	Achieve at least 70% on overall accuracy	Achieve at least 60% on overall accuracy	Achieve less than 60% on overall accuracy
		Countvec (15%)	Achieve at least 90% on overall accuracy	Achieve at least 80% on overall accuracy	Achieve at least 70% on overall accuracy	Achieve at least 60% on overall accuracy	Achieve less than 60% on overall accuracy
	<b>Methodology (35%)</b>	Data processing techniques(20 %)	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(15% )	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.
	<b>Documentation (15%)</b>	Report structure and discussions(10%)	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly. The report can be improved.	The report has some sectioning but it is not well organised. The explanation of the report is limited	The report has some sectioning but it is not well organised. The explanation of the report is poor.	The report has poor/no sectioning and it is not well organised. The explanation of the report is poor.
		Code structure and documentation (5%)	Code is well-structured, easy to read, and follows best practices. Code is thoroughly documented with clear and concise comments.	Code is generally readable but has structural issues. Deviations from best practices are noticeable. Code is well-documented with minor gaps in comments.	Code is somewhat readable but lacks structure. Deviations from best practices are noticeable. Code has minimal documentation and comments.	Code is somewhat readable but poorly structured. Deviates significantly from best practices. Code has few comments and	Code is poorly structured and hard to read. Does not follow best practices. Code lacks documentation and comments

						lacks proper documentation.	
<b>Task 3 (12/35)</b>	<b>Exploratory Data Analysis</b>						
	<b>Methodology(85%)</b>	Insight Relevance(30%)	Provides 4 or more high-quality insights(and 2 or more used both the auxiliary data and the input dataset from task1). Insights are highly relevant and provide significant value. Strong implications for further research or business.	Provides 3 or more high-quality insights.Insights are relevant and provide good value. Useful for further research or business.	Provides at least 2 high-quality insights. Insights are somewhat relevant and provide moderate value.	Provides at least 2 high-quality insights. Insights are minimally relevant and provide limited value.	Insights are not relevant and provide little to no value.
		Support with Data(15%)	Insights are well-supported by data with strong evidence.	Insights are supported by data with good evidence.	Insights are somewhat supported by data but lack strong evidence.	Insights have minimal support from data and weak evidence.	Insights are not supported by data and lack evidence.
		Initial Data Exploration (30%)	Thorough initial exploration of data, including summary statistics and visualisation.	Good initial exploration of data with minor gaps in summary statistics or visualisation.	Adequate initial exploration of data but lacking in detail or depth.	Basic initial exploration of data with significant gaps in summary statistics or visualisation.	Poor initial exploration of data with minimal or no summary statistics or visualisation.
		Justification and Explanation(10%)	Provides thorough justification and detailed explanation for each insight, linking it to the data analysis.	Provides good justification and explanation for each insight, with minor gaps in linking to the data analysis.	Provides basic justification and explanation for each insight, but some links to the data analysis are weak.	Provides minimal justification and explanation for insights, with poor links to the data analysis.	Provides no or poor justification and explanation for insights, with no links to the data analysis.
	<b>Documentation (15%)</b>	Report structure and discussions(10%)	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly. The report can be improved.	The report has some sectioning but it is not well organised. The explanation of the report is limited	The report has some sectioning but it is not well organised. The explanation of the report is poor.	The report has poor/no sectioning and it is not well organised. The explanation of the report is poor.

		Code structure and documentation (5%)	Code is well-structured, easy to read, and follows best practices. Code is thoroughly documented with clear and concise comments.	Code is generally readable but has structural issues. Deviations from best practices are noticeable. Code is well-documented with minor gaps in comments.	Code is somewhat readable but lacks structure. Deviations from best practices are noticeable. Code has minimal documentation and comments.	Code is somewhat readable but poorly structured. Deviates significantly from best practices. Code has few comments and lacks proper documentation.	Code is poorly structured and hard to read. Does not follow best practices. Code lacks documentation and comments
<b>Task 4 (2/35)</b>	<b>Video presentation</b>						
	<b>Presentation Clarity(50%)</b>		Presentation is between 5-8 minutes. Presentation is exceptionally clear, well-organised, and easy to follow.	Presentation is clear and organised with minor areas for improvement.	Presentation is generally clear but has some organisation issues.	Presentation is somewhat clear but lacks organisation and smooth flow.	Presentation is unclear, poorly organised, and hard to follow.
	<b>Content Coverage(45%)</b>		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.
	<b>Delivery and Engagement(5%)</b>		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.
<b>Task 5 (1/35)</b>	<b>Development History</b>						
	<b>Task 1 Record (1/3)</b>		The report provides a detailed timeline highlighting at least three significant time points in the development of your assignment and includes screenshots or snapshots of different versions of your assignment at each time point. If applicable, the report contains sufficient information on the collaborative effort.				
	<b>Task 2 Record (1/3)</b>		The report provides a detailed timeline highlighting at least three significant time points in the development of your assignment and includes screenshots or snapshots of different versions of your assignment at each time point. If applicable, the report contains sufficient information on the collaborative effort.				
	<b>Task 3 Record (1/3)</b>		The report provides a detailed timeline highlighting at least three significant time points in the development of your assignment and includes screenshots or snapshots of different versions of your assignment at each time point. If applicable, the report contains sufficient information on the collaborative effort.				