

The Rubric for A2

Task 1 & Task 2 (95%)

		HD	D	C	P	F
Complete Task 3		This is the HURDLE of A2.				
Correct structure for the output files as specified in the specs		It is a must. Note that the output files should be parsable and also be in the pre-defined structure and with the same column names. Unparsable outputs and the output files that have any deviations from the specification, will receive zero marks for the output.				
Output (55%)	Dirty Data (30%)	80% or more of the errors are fixed accurately.	70% - 79% of the errors are fixed accurately.	60% - 69% of the errors are fixed accurately	50% - 59% of the errors are fixed accurately	0% - 49% of the errors are fixed accurately
	Outliers (10%)	80% or more of the outliers are removed.	70% - 79% of the outliers are removed.	60% - 69% of the outliers are removed.	50% - 59% of the outliers are removed.	0% - 49% of the outliers are removed.
	Missing Values (15%)	80% or more of the missing values are imputed correctly.	70% - 79% of the missing values are imputed correctly.	60% - 69% of the missing values are imputed correctly.	50% - 59% of the missing values are imputed correctly.	0% - 49% of the missing values are imputed correctly.
Task 1 Methodology (15%)		The report demonstrated a good solution in a proper way. No error or only minor errors are found in the methodology. The methodology consists of all required steps for the tasks, and produces output with HD scores.	The methodology consists of all required steps for the tasks, and produces output with above C scores.	The report has provided a fair solution. Three key steps are missing or lead to incorrect output (above P).	The report has provided a poor solution. Four key steps 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..
Task 2 Data Reshape (15%) Exploratory Task of Data Reshaping		Different methods for data reshaping are well analyzed. The effects of using different data-reshaping methods are shown in detail. The student uses different attributes to evaluate the performance of linear regression models and gives details explanations. The student has made a convincing recommendation to the readers based on the observation in the analysis.	Different methods for data reshaping are reasonably analyzed. The effects of using different data-reshaping methods have been demonstrated. The student uses different attributes to evaluate the performance of linear regression models and gives a detailed explanation. The student has made a	Different methods for data reshaping are fairly analyzed. The effects of using different data-reshaping methods have been demonstrated. The student uses different attributes to evaluate the performance of linear regression models and gives little explanation. The student has made a recommendation	Different methods for data reshaping are poorly analyzed. The effects of using different data-reshaping methods have been demonstrated. The student builds a simple linear regression model without much consideration and explanation. The student then rushes to a recommendation to the readers.	Different methods for data reshaping are badly analyzed. The effects of using different data-reshaping methods have not been demonstrated. The student builds a simple linear regression model without any explanation. The student then rushes to a

		reasonable recommendation to the readers based on the observation in the analysis.	to the readers based on the observation in the analysis.		recommendation to the readers.
Documentation (10%)	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained properly and the code is well commented.	The report has proper sections and subsections (e.g. introduction, methodology, conclusion,...). The methodology is explained fairly and the code is fairly commented. The report can be improved.	The report has some sectioning but it is not well organized. The explanation of the report is limited and the code is well commented.	The report has some sectioning but it is not well organized. The explanation of the report is poor and the code is well commented.	The report has poor/no sectioning and it is not well organized. The explanation of the report is poor and the code is poorly commented.

Note: Both non-fixed errors and newly introduced errors would be penalized (i.e. if while attempting to fix one of the anomalies - dirty data, missing value, and outliers- you introduced another error, you will lose marks accordingly)

Task 3 (5% + Hurdle)

	Satisfied (1)	Not Satisfied (0)
Time Limit (Approx. 5 min)	5-10 minutes	Less than 2 minutes, or over 10 minutes
Attendance	Show up	Not show up
Contribution	Both members talk about their A2	Not both member talk about A2
Coverage of A2	At least cover majority of task 1	Less than 70% of A1 was covered.
Question Answered	Able to answer questions from TA	Fail to answer questions from TA

The Rubric for A2 – Individual Score Calculation

The individual score for each group member will be calculated as:

$$\text{Individual Score} = \text{Score of A2} * \min\left(1, \frac{\text{contribution\%}}{50\%}\right)$$

This formula is designed to determine the individual score of each member within a group, based on their contribution percentage to the group's score.

Each individual's score is calculated by multiplying the group score by their contribution percentage, ensuring that their score is directly proportional to their contribution. However, to achieve fairness and consistency, the formula also includes a restriction that prevents an individual's score from exceeding the group score.

For instance, if the group score is 100 and both members contribute equally at 50%, both members will receive the same individual score of $100 \times \min(1, 50/50) = 100$. If a member's contribution is greater than 50%, their score will still be capped at the group score.