

## The Rubric for A2

		HD	D	C	P	F
	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 (60%)	Dirty Data (30%)	90% or more of the errors are fixed accurately.	75% - 89% of the errors are fixed accurately.	60% - 74% of the errors are fixed accurately	50% - 59% of the errors are fixed accurately	0% - 49% of the errors are fixed accurately
	Outliers (15%)	90% or more of the outliers are removed.	75% - 89% of the outliers are removed.	60% - 74% of the outliers are removed.	50% - 59% of the outliers are removed.	0% - 49% of the outliers are removed.
	Missing Values (15%)	90% or more of the missing values are imputed correctly.	75% - 89% of the missing values are imputed correctly.	60% - 74% 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.
Methodology (25%)		The report has 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 report has demonstrated a reasonable solution. Two key step is missing or leads to the incorrect output.	The report has provided a fair solution. Three key steps are missing or lead to incorrect output.	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..
Documentation (15%)		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)