# The Rubric for A2

## Task 1 & Task 2 (30/40)

| | | HD | D | C | P | F |
|---|---|---|---|---|---|---|
| **Complete Task 3** | | This is the HURDLE of A2, failure to complete will result in ZERO 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%) | 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 (10%) | 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. |
| **Task 1 Methodology (10%)** | | 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 (25%)**<br><br>Exploratory Task of Data Reshaping | | Different methods for data reshaping are well analysed. 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 analysed. 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 reasonable | Different methods for data reshaping are fairly analysed. 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 to the readers | Different methods for data reshaping are poorly analysed. 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 analysed. The effects of using different data-reshapin g methods have not been demonstrated. The student builds a simple linear regression model without any explanation. The student then rushes to a recommendati |

| | | | | | |
|---|---|---|---|---|---|
| | | recommendation to the readers based on the observation in the analysis. | based on the observation in the analysis. | | on 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 organised. The explanation of the report is limited and the code is well commented. | The report has some sectioning but it is not well organised. The explanation of the report is poor and the code is well commented. | The report has poor/no sectioning and it is not well organised. 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 (10/40 + Hurdle)

| | **Yes (2)** | **Partial(1)** | **No (0)** |
|---|---|---|---|
| **Able to Answer Question 1** | Able to answer the question correctly | Answer (partially correct) | Fail to answer question/ Answer is not correct |
| **Able to Answer Question 2** | Able to answer the question correctly | Answer (partially correct) | Fail to answer question/ Answer is not correct |
| **Able to Answer Question 3** | Able to answer the question correctly | Answer (partially correct) | Fail to answer question/ Answer is not correct |
| **Able to Answer Question 4** | Able to answer the question correctly | Answer (partially correct) | Fail to answer question/ Answer is not correct |
| **Able to Answer Question 5** | Able to answer the question correctly | Answer (partially correct) | Fail to answer question/ Answer is not correct |