**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 15 March 2024 |
| Team ID | **739687** |
| Project Title | Work force retention system |
| Maximum Marks | 2 Marks |

**Data Quality Report:**

The Data Quality Report will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data

discrepancies.



**Data Quality Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data**  **Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
|  | Missing values in the 'Satisfactory level', ‘last evaluation’, |  |  |
| Kaggle  Dataset | ‘Number of projects’,  ‘Average monthly hours’,  ‘Time spend company’,  ‘Work accident’, and ‘salary’ columns. | Moderate | Use mean/median imputation. |
| Kaggle  Dataset | Categorical data in the dataset | Moderate | Encoding has to be done in the data. |