

|  |  |  |
| --- | --- | --- |
|  | **Data Collection and Preprocessing Phase** | |
|  |  |  |
| Date |  | 15 March 2024 |
|  |  |  |
| Team ID |  | PNT2022TMID124356 |
|  |  |  |
| Project Title |  | SmartLender - Applicant Credibility |
|  |  | Prediction for Loan Approval |
|  |  |  |
| 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 |  |  |  |
|  | 'Gender', ‘Married’, |  |  |  |
| Kaggle | ‘Dependents’, |  | Use mean/median |  |
| ‘Self\_Employed’, | Moderate |  |
| Dataset | imputation. |  |
| ‘LoanAmount’, |  |  |
|  |  |  |  |
|  | ‘Loan\_Amount\_Term’, and |  |  |  |
|  | ‘Credit\_History’ columns. |  |  |  |
|  |  |  |  |  |
| Kaggle | Categorical data in the dataset | Moderate | Encoding has to be done in |  |
| Dataset | the data. |  |
|  |  |  |
|  |  |  |  |  |