**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 9 July 2024 |
| Team ID | 739738 |
| Project Title | Leveraging Machine Learning For GDP Per Capita Prediction |
| Maximum Marks | 2 Marks |

**Data Quality Report Template**

The Data Quality Report Template 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 Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | Missing Values | High | Impute missing values using mean or median for numerical features. For categorical features, use mode. |
| Dataset | Outliers | Moderate | Use statistical methods like Z-score or IQR to detect and handle outliers, either by removing or transforming them. |
| Dataset | Duplicate Records | Low | Remove duplicate records from the dataset to ensure each instance is unique |
| Dataset | Data Currency | High | Ensure the dataset is up-to-date by verifying the date range and considering additional data collection if necessary. |
| Dataset | Inconsistent Records | Moderate | Standardize date formats and numerical representations across columns using data preprocessing techniques. |