**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| **Date** | 11 March 2025 |
| **Team ID** | 740052 |
| **Project Title** | AI-Based Intelligent Insight Extractor |
| **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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | | **Component** | **Insight Extracted** | **Severity** | **Resolution Plan** | |  |  |  |  | | Data Profiler | Categorical features detected | Moderate | Apply encoding: one-hot or label encoding | | Document Analyzer | Extracted metadata and manual annotations from project documentation | Low | Integrate document insights into preprocessing workflow | | Recommendation Engine | Suggested preprocessing strategy for modelling | High | Auto-generate a pipeline: imputation → encoding → scaling | | Insight Visualizer | Highlights severity and transformation impact | Low | Display visual report for analyst review | | Automation Engine | Identified repeatable data quality checks | High | Schedule routine data checks with Airflow or similar tools | |