**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 18 June 2024 |
| Team ID | **739990** |
| Project Title | Auto Insurance Fraud Detection |
| 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** |
| Kaggle  Dataset | Missing values in the  ‘'collision\_type',  ’property\_damage’,  ‘police\_report\_available’,  ‘age’,’total\_claim\_amount’,  ‘authorities\_contacted’  columns. | Moderate | Use mean/median imputation. |
| Kaggle  Dataset | Categorical data in the dataset | Moderate | Encoding has to be done in the data. |