**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 17 April 2024 |
| Team ID | 738220 |
| Project Title | Walmart Sales Analysis for Retail Industry with Machine Learning |
| 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 Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Kaggle | Missing Values in the columns:  ‘MarkDown1’,  ‘MarkDown2’  ‘MarkDown3’,  ‘MarkDown4’ | Low | Filled the missing values with the 0s. |
| Kaggle | Negative values in the ‘Temperature’ and ‘Weekly\_Sales’ columns | Low | Used Conditions to filter out values greater than or equal to 0. |
| Kaggle | SettingWithCopyWarning | Moderate | Solved by using loc accessor |
| Kaggle | Issue in making Heatmap | Moderate | Solved by encoding all the categorical columns into numbers. |