** **

**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 April 2024 |
| Team ID | 739942 |
| Project Title | Freedom of the World Classification |
| 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.

This report outlines the data quality issues identified in the Freedom of the World classification data, along with their severity and proposed resolution plans. The goal is to ensure data integrity, accuracy, and reliability.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | | **Severity** | **Resolution Plan** |
| Kaggle Dataset | Booking Status are present in the dataset | | Low | Dropped Booking Status using drop() function |
| Kaggle Dataset | Missing values are present in the dataset | | Low | Removed the missing values from dataset by using the following code:  test\_data.isna().sum()  train\_data.isna().sum() |
| Kaggle Dataset | | Outliers are present in some of the columns of dataset | Medium | Identified outliers by drawing boxplot and removed them from upper bound and lower bound. |
|  | | | | |