** **

**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 April 2024 |
| Team ID | 740661 |
| Project Title | RESERVATION CANCELLATION 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** |
| 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. |
|  | | | | |
|  |