** **

**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 April 2024 |
| Team ID | 740140 |
| Project Title | Crystal Ball Analysis: Projecting share prices of the leading GPU titans |
| 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 | |  | | --- | | Missing Values |  |  | | --- | |  | | | High | Implement imputation strategies (mean, median, etc.) |
| Kaggle Dataset | Outliers | | Medium | Evaluate impact on analysis; consider filtering or transformation |
| Kaggle Dataset | | Inconsistent Formatting | Low | Standardize formats (dates, currencies, etc.) |
| Kaggle Dataset | | Data Inaccuracies (e.g., duplicates, incorrect values) | Medium | |  | | --- | |  |  |  | | --- | | Conduct data cleaning procedures;  verify against primary sources | |
| Kaggle Dataset | | Limited Historical Data | Low | Augment dataset with additional sources or proxy variables |