Data Collection and Preprocessing Phase

|  |  |
| --- | --- |
| Date | 19 March 2025 |
| Team ID | LTVIP2025TMID60925 |
| Project Title | Visualizing Housing Market Trends: An Analysis of Sale Prices and Features using Tableau |
| Maximum Marks | 3 Marks |

**Data Quality Report Template**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Transformed Housing Data 2 | * Missing values in key attributes. * Duplicate records. * Outliers affecting analysis. * Incorrect data types or formats. * Inconsistent data entry. | Moderate | * Filling missing values using mean/median/mode. * Removing or correcting duplicates. * Normalizing or transforming data types. * Handling outliers using statistical techniques. |
| Transformed  Housing Data 2 | Duplicate records found (11 instances). | Moderate | Remove duplicate records using data cleaning techniques. |
| Transformed  Housing Data 2 | No of Times Visited' column contains only zero values, which may indicate missing or incorrect data. | High | Verify data source or remove column if it lacks meaningful information. |

|  |  |  |  |
| --- | --- | --- | --- |
| Transformed  Housing Data 2 | Zipcode representation is in multiple binary columns instead of a single categorical variable. | Moderate | Convert multiple binary columns into a single categorical 'Zipcode Group' column for better efficiency. |
| Transformed  Housing Data 2 | Potential outliers in 'Sale Price' and 'Lot Area' columns. | High | Use statistical methods to identify and handle outliers (e.g., IQR method). |