**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 30 July 2025 |
| ID | SWUID20250188226 |
| Project Title | Global Food Production Analysis (1961-2023) |
| Maximum Marks | 3 Marks |

**Data Quality Report Table**

| **Data Sorce** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| --- | --- | --- | --- |
| Dataset | Missing values in crop production columns for specific countries and years | Moderate | Used data imputation techniques such as forward fill and interpolation for continuous year series |
| Dataset | Inconsistent naming conventions (e.g., “USA” vs “United States”, “Maize” vs “Corn”) | Moderate | Standardized using mapping dictionaries and applied consistent country/crop names |
| Dataset | Presence of zero or extremely low values for major crops in major producing countries | High | Cross-verified with external sources (FAO/World Bank) and replaced incorrect outliers with median values |
| Dataset | Duplicate rows with identical year, country, and crop data | Low | Removed duplicates using pandas drop\_duplicates() function |