**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 01 October 2025 |
| Team ID | SWUID20250215169 |
| Project Title | Global Food Production Trends and Analysis:  A Comprehensive Study from 1961 to 2023 Using Power BI |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The dataset, titled “World Food Production (FAO)”, contains annual food production data for various crops across countries and regions from 1961 to 2023. Key fields include Entity (Country/Region), Crop, Year, and Production (tonnes). The source was FAO’s open agricultural statistics database. |
| Data Cleaning | Using Power Query, missing values and duplicates were identified and removed. Inconsistent or null entries in Production (tonnes) were corrected. Crop names and region labels were standardized to ensure consistency across records. |
| Data Transformation | The dataset was unpivoted so that each year and its production value became separate rows, enabling dynamic time-series analysis in Power BI. Further, Power Query transformations were applied to filter data between 1961–2023, and new calculated columns were created such as Decade, Entity Type (Country/Region), and Crop Category. |
| Data Type Conversion | Corrected data types to ensure smooth aggregation — Year converted to whole number, Production (tonnes) to decimal, and Entity and Crop to text. This ensured compatibility for visuals and DAX measures. |
| Column Splitting and Merging | Split or merge columns as needed. |
| Save Processed Data | After cleaning, unpivoting, and transforming, the processed dataset was saved as ‘world\_food\_production\_preprocessed’ and loaded into Power BI for dashboard and report development. |