**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 June 2025 |
| Team ID | xxxxxx |
| Project Title | Global Food Production Trends and Analysis (1961–2023) using power bi |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing Template**

The project uses the World Food Production dataset (1961–2023), addressing issues like missing values, duplicates, and inconsistent naming. Resolution steps ensure clean, accurate, and reliable data for analysis and visualization.

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | |  | | --- | | Dataset contains global food production records from 1961–2023 with key fields such as Country, Commodity, Year, Production (tonnes), Area Harvested (hectares), and Yield (tonnes/hectare). |  |  | | --- | |  | |
| Data Cleaning | |  |  |  | | --- | --- | --- | | |  | | --- | | Some missing values exist for smaller countries/commodities in certain years. Inconsistent country/commodity naming identified. Cleaning involves standardization and interpolation where necessary. |  |  | | --- | |  | |  |  | | --- | |  | |
| Data Transformation | |  |  |  | | --- | --- | --- | | |  | | --- | | Create derived columns such as Per-Capita Production, CAGR, Global Share (%), and decade groupings for analysis. |  |  | | --- | |  | |  |  | | --- | |  | |
| Data Type Conversion | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  | | --- | --- | --- | --- | --- | | |  |  |  | | --- | --- | --- | | |  | | --- | | Ensure numeric columns (Production, Area, Yield) are floats/integers, while Country, Commodity, and Year are categorical/date types. |  |  | | --- | |  | |  |  | | --- | |  | |  |  | | --- | |  | |  |  | | --- | |  | |
| Column Splitting and Merging | |  | | --- | | No direct combined columns, but merging Year into Decade buckets and grouping commodities into categories (e.g., Cereals, Oilseeds, Pulses) is required. |  |  | | --- | |  | |
| Data Modeling | |  |  |  | | --- | --- | --- | | |  | | --- | | Organize into a star schema with FactProduction linked to DimCountry, DimCommodity, and DimDate tables for efficient dashboarding in Power BI. |  |  | | --- | |  | |  |  | | --- | |  | |
| Save Processed Data | After preprocessing, save as processed\_food\_production.csv or load into Power BI for trend analysis and dashboard creation. |