**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 23 July 2024 |
| Team ID | - |
| Project Title | Global Food Production Trends and Analysis |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The dataset contains production data of major agricultural commodities from 1961 to 2023, across 226 countries or regions. It includes annual production figures (in tonnes) for items like rice, wheat, maize, fruits, and beverages. |
| Data Cleaning | Standardized column names by removing whitespaces. Checked for missing values and outliers. No nulls were found, but zero values were retained after context review. Duplicate records were not found in the dataset. |
| Data Transformation | Use of Power Query for filtering, sorting and creating calculated columns. |
| Data Type Conversion | Ensured numerical columns were of float type and corrected the 'Year' to integer format. Confirmed all categorical fields such as 'Entity' are in string format. |
| Column Splitting and Merging | Merged related crop categories into broader groups like 'Cereals' and 'Fruits'. Split complex headers like 'Coffee, green Production ( tonnes)' into simpler forms during preprocessing for clarity. |
| Data Modeling | Data was modeled in Power BI by treating each crop as a fact field and 'Entity' and 'Year' as dimensions. Relationships were defined using normalized entity and year tables where applicable. |
| Save Processed Data | The processed data was saved in Power BI's data model for use in reports and dashboards. |