

Data Collection and Preprocessing Phase

Date	30 September 2025
Team ID	SWUID20250181744
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

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	<p>The dataset contains global food production records from 1961 to 2023, capturing key agricultural commodities such as rice, wheat, maize, apples, avocados, bananas, oranges, grapes, tea, and green coffee. Each record includes the Entity (country/region), a unique code, and the annual production quantity (in tonnes). The data provides a comprehensive overview of global agricultural trends across more than six decades, serving as the foundation for visualization and trend analysis in Power BI.</p>
Data Cleaning	<p>During cleaning, the dataset was inspected for missing values, duplicates, and inconsistencies.</p> <ul style="list-style-type: none">• Missing production values were replaced with zeros or interpolated based on previous years to maintain trend continuity.• Duplicate records (same entity, year, and commodity) were identified and removed using Power Query's Remove Duplicates feature.• Inconsistent entity names (e.g., "USA" vs "United States") were standardized for uniformity.• Outliers with unrealistically high or low production figures were flagged for review.

Section	Description
Data Transformation	<p>The Power Query Editor in Power BI was used extensively for data shaping.</p> <ul style="list-style-type: none"> • Unnecessary columns were filtered out. • Sorting and pivoting were applied to reorganize data by year and commodity. • New calculated columns were created to compute total production per year and percentage contributions by commodity and region.
Data Type Conversion	<p>Each column was verified and converted to the correct data type:</p> <ul style="list-style-type: none"> • <i>Entity</i> → Text • <i>Year</i> → Whole Number • Production columns → Whole Number (tonnes). <p>This ensured that Power BI could perform accurate aggregations and time-series analyses.</p>
Column Splitting and Merging	<p>Column operations were performed to enhance usability.</p> <ul style="list-style-type: none"> • The Entity and Code fields were retained separately to allow for both textual and coded references. • No major splitting was required, but some combined columns were separated for clarity. • Aggregated production metrics were created by merging related production columns using DAX calculations.
Data Modeling	<p>A star schema was created in Power BI for efficient data analysis.</p> <ul style="list-style-type: none"> • The Fact Table contains annual production values for all commodities. • Dimension Tables include Entity (Country/Region), Commodity, and Date (Year). • Relationships were established between these tables using primary keys such as Entity and Year. • DAX measures were developed to calculate KPIs like total production, year-over-year growth, and commodity share.

Section	Description
Save Processed Data	The cleaned and transformed dataset was saved as a Power BI Data Model (.pbix) file and also exported as a cleaned CSV for future analysis. This ensures reproducibility, allowing updates or new visualizations to be built easily without repeating the preprocessing steps.