Data Collection and Preprocessing Phase

Date	14 Sept 2025
Team ID	SWUID20250185946
Project Title	World Food Production Analysis 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, sourced from Kaggle, spans global food production data from 1961 to 2021, featuring structured CSV/Excel formats with attributes such as year, crop type, harvested area, production volume, and yield.
Data Cleaning	Missing values were addressed using averaging and interpolation techniques. Duplicate entries were eliminated, and naming inconsistencies in countries and crops were standardized to ensure data accuracy and consistency.
Data Transformation	Power Query was utilized to filter out non-essential attributes, organize data by year and country, generate calculated columns such as yield (production ÷ area), and create pivot tables for streamlined visualization.
Data Type Conversion	Year values were converted to whole numbers, numeric fields like Production, Yield, and Area were formatted as decimals, and Country/Crop names were standardized as text to ensure accurate analysis.
Column Splitting and Merging	Combined fields such as '1961F' were split into separate components like Year and Type, while key columns such as Country and Crop were merged to create unique identifiers for more accurate analysis.

Data Modeling	Established relationships between the fact table (production data) and dimension tables (year, crop, country) to enable structured analysis. Developed DAX measures such as Total Production, Total Area, and Average Yield to support insightful reporting and performance evaluation.
Save Processed Data	The finalized dataset and Power BI model (.pbix) were saved, with backups of both raw and processed data preserved for future reference and analysis.