

Data Collection and Preprocessing Phase

Date	25-12-2025
Team ID	
Project Title	Global Food Production Analysis (1961–2023)
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

This phase focuses on identifying reliable data sources and planning the collection and preprocessing of raw data to ensure accuracy, consistency, and suitability for analysis. Proper data collection and preprocessing are essential to generate meaningful insights and support informed decision-making.

Data Collection Plan Template

Section	Description
Project Overview	This project aims to analyze global food production trends from 1961 to 2023 by collecting historical agricultural data and transforming it into meaningful visual insights using Power BI.
Data Collection Plan	Data is collected from publicly available and reliable global agricultural datasets that provide yearly production statistics for major food crops across different countries and regions.
Raw Data Sources Identified	Raw datasets containing crop-wise and year-wise production values in tonnes are identified and downloaded in CSV/Excel formats for further preprocessing and analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Global food production data containing crop-wise production values from 1961 to 2023	Public agricultural data repository (e.g., FAO / Kaggle)	CSV	50 MB	Public
Dataset 2	Regional and country-level agricultural production statistics.	Public data source	Excel	30 MB	Public
Dataset 3	Crop category reference dataset for classification	Internal prepared dataset	CSV	5 MB	Private (Project Use)