**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 18 August 2025 |
| Name | Mohammed Farhan |
| Project Title | India Agriculture Crop Production Analysis (1997–2021) Using Tableau |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | Analyze India’s crop production across states from 1997 to 2021 and surface long‑term trends, regional variations, and crop performance using interactive Tableau dashboards and Tableau Stories. Data is curated from Government open data and validated against official publications. |
| Data Collection Plan | • Identify primary sources: Government of India Open Data (DES/MoAFW) APY datasets. • Download annual/state/crop CSVs (1997–2021); document schema and units. • Create lookup mappings for crop synonyms and legacy state names. • Version datasets in GitHub; maintain raw → cleaned → modeled folders. • Validate aggregates with ‘Agricultural Statistics at a Glance’. • Store working copies in MySQL and export final CSVs for Tableau. |
| Raw Data Sources Identified | Primary: Directorate of Economics & Statistics (DES) APY datasets (1997–2021). Secondary: MoAFW ‘Agricultural Statistics at a Glance’; FAOSTAT India aggregates. Working copies: curated CSVs and lookup tables stored in the project’s GitHub repo. |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| GoI DES – APY (1997–2021) | State × Crop annual production statistics; primary dataset. | https://data.gov.in/ | CSV | ~150–200 MB | Public |
| MoAFW – Agricultural Statistics at a Glance | Official yearly compendium used to validate totals and trends. | https://agricoop.gov.in/ | PDF/XLSX | Varies | Public |
|  |  |  |  |  |  |
| FAOSTAT – India Crops | International aggregates for external validation. | https://www.fao.org/faostat/ | CSV | ~20–30 MB | Public |
| GitHub Repo – Curated Dataset | Cleaned & standardized CSVs used in Tableau workbook. | https://github.com/mohammedfarhan14/India-Agriculture-crop-production-analysis/tree/master/Dataset | CSV | ~50–120 MB | Public |
|  |  |  |  |  |  |
| Lookup Tables – State & Crop Mapping | Mapping dictionaries for legacy state names and crop synonyms. | GitHub Dataset/lookups folder | CSV | <1 MB | Public |
| MySQL Workbench – Project DB | Local relational store for joins, checks, and exports. | Local Environment | SQL/CSV Exports | Project-scale | Private |

**Raw Data Sources Template**