**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 | 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 | Dataset: India Agriculture Crop Production (1997–2021) Source: Government of India Open Data Portal / Directorate of Economics and Statistics Format: CSV files containing structured tabular data Size: ~200 MB Fields: Year, State, Crop, Area, Production, Yield Scope: Covers multiple states and crops over 25 years |
| Data Cleaning | • Removed duplicate records caused by multiple uploads • Standardized crop names (e.g., “Ground nut” → “Groundnut”) • Corrected inconsistent state naming (e.g., “Orissa” → “Odisha”) • Handled missing values:  - Production/area/yield: median imputation  - State/Crop categories: validated via external sources |
| Data Transformation | • Applied filtering for major crops (Rice, Wheat, Maize, Pulses, Sugarcane, Cotton) • Sorting applied to rank states/crops by production and yield • Pivot tables created for Year vs. Crop and State vs. Crop production • Calculated fields:  - Yield = Production ÷ Area  - Growth % = ((Current Year – Previous Year) ÷ Previous Year) × 100 |
| Data Type Conversion | • Year → Integer • Production, Area → Float/Decimal • State, Crop → String (Categorical) |
| Column Splitting and Merging | • Split combined fields (e.g., Andhra Pradesh (before bifurcation) → Andhra Pradesh & Telangana) • Merged columns to create consolidated indicators like Total Production |
| Data Modeling | • Relationships defined:  - Year ↔ Production (time-series trends)  - State ↔ Crop (regional comparisons) • Tableau data model built with:  - Fact table: production data  - Dimension tables: state, crop, year |
| Save Processed Data | • Final cleaned dataset saved as CSV (Processed\_Agri\_Data.csv) • Uploaded to GitHub repository for transparency & version control • Tableau packaged workbook (.twbx) saved with embedded processed dataset |