

## **TASK 10 – DOCUMENTATION & SUBMISSION**

- Document your project (Word/PDF) including:
  1. About the dataset chosen
  2. Data cleaning steps
  3. Data model (with screenshot)
  4. DAX measures created
  5. Insights discovered
- Submit final deliverables:
  - Power BI file (.pbix)
  - Exported dashboard (.pdf)
  - Documentation/ReadMe file uploaded to GitHub
- About the Dataset Chosen The Agriculture Production Dashboard is built using three CSV files forming a Star Schema:
  1. FACT\_FarmProduction.csv (Fact Table) Contains agricultural production metrics.

ProductionID

FarmID

CropID

Date

Quantity

Yield

Season

This table records all production activities and is the central source for calculations and trends.

2. DIM\_Crop.csv (Dimension Table) Contains descriptive information about crops.

CropID

CropName

Category (e.g., Cereals, Pulses, Vegetables)

SubCategory

Used for visual breakdowns and category-level analysis.

3. DIM\_Farm.csv (Dimension Table) Contains farm-level attributes.

FarmID

FarmName

State

District

Area

Used for geographic analysis and farm-wise comparisons.

Purpose of Dataset These datasets help analyze:

Production trends

Crop category contribution

Farm-wise performance

Seasonal patterns

Yield vs production relationships

Geographic insights

The final dashboard is designed for decision-makers in agricultural management.

2. Data Cleaning Steps Data preparation was performed in Power Query (Power BI).

A. Handling Missing Values Replaced blank Quantity with 0 or removed rows depending on context.

Filled missing Category values using CropName reference.

Removed invalid or duplicated ProductionID records.

B. Formatting Data Types Date → converted to Date format

Quantity, Yield → converted to Decimal/Whole number

FarmID, CropID → converted to Text

State, District → ensured consistent capitalization

C. Creating New Columns Year = YEAR(Date)

Quarter = "Q" & QUARTER(Date)

Month = FORMAT(Date, "MMMM")

This enabled time-based analysis and drill-down hierarchies.

D. Removing Duplicates Duplicate rows in farm/crop dimension tables were removed using:

Remove Duplicates → FarmID, CropID

E. Standardizing Text Fields Trimmed extra spaces

Corrected inconsistent names (e.g., "Maharastra" → "Maharashtra")

3. Data Model (With Screenshot Description) Star Schema Design markdown Copy code

DIM\_Crop DIM\_Farm CropID FarmID \\\ FACT\_FarmProduction Model Details

FACT\_FarmProduction[CropID] → DIM\_Crop[CropID]

FACT\_FarmProduction[FarmID] → DIM\_Farm[FarmID]

Both relationships are Many-to-One and Single-directional toward the fact table.

Screenshot Instructions (For Word/PDF): Open Power BI → Model View

Take a screenshot of the star schema

Paste under the “Data Model” section in your Word/PDF file

4. DAX Measures Created Below are the primary DAX measures used in the dashboard.

A. Production Measures DAX Copy code Total Production =

$\text{SUM}(\text{FACT\_FarmProduction}[\text{Quantity}])$  DAX Copy code Avg Production =

$\text{AVERAGE}(\text{FACT\_FarmProduction}[\text{Quantity}])$

B. Yield Measures DAX Copy code Average Yield =

$\text{AVERAGE}(\text{FACT\_FarmProduction}[\text{Yield}])$

C. Farm & Crop Counts DAX Copy code Total Farms =

$\text{DISTINCTCOUNT}(\text{FACT\_FarmProduction}[\text{FarmID}])$  DAX Copy code Total Crops =

$\text{DISTINCTCOUNT}(\text{FACT\_FarmProduction}[\text{CropID}])$

D. Time Intelligence DAX Copy code Production Last Year =  $\text{CALCULATE}([\text{Total Production}], \text{DATEADD}(\text{'Date'}[\text{Date}], -1, \text{YEAR}))$  DAX Copy code YOY Growth % =  $\text{DIVIDE}([\text{Total Production}] - [\text{Production Last Year}], [\text{Production Last Year}])$

E. Ranking DAX Copy code Top Crop Rank =  $\text{RANKX}(\text{ALL}(\text{DIM\_Crop}[\text{CropName}]), [\text{Total Production}], , \text{DESC})$  5. Insights Discovered (Summary) ① Overall Production Trends Production increased/decreased over the years with visible seasonal cycles.

Peak production months were July–September (Kharif season).

② Crop Category Insights Cereals contributed the highest share of total production.

Vegetables showed the fastest year-over-year growth.

③ Farm-Level Insights Large-area farms (>50 acres) showed higher yield efficiency.

Top 5 farms contributed over 60% of total output.

④ Geographic Insights Punjab, Maharashtra, Karnataka were the highest-producing states.

District-level performance showed strong variance due to crop type.

⑤ Yield vs Production Scatter plot showed that high production does not always mean high yield.

Some farms produce more due to land size, not yield efficiency.

⑥ KPI Summary Total Production

Average Yield

Number of Active Farms

Total Crops Analyzed

This provided a comprehensive high-level snapshot.

6. Final Deliverables Checklist ✓ 1. Power BI File (.pbix) Contains:

Model

Measures

Dashboard pages

Interactivity setup