# Crop Data Insights Report

**Data Preparation Process**

1. **Source Data**:

   o Initially read from crops_data.csv (2010 – 2017 Data), which contained **Row based crop statistics**.

2. **Reshaping the Data**:

   o Applied melt() to transform the data from **row format** to **column format** where each crop had:

     ▪ **Area** (in 1000 ha), later converted into ha in power BI

     ▪ **Production** (in 1000 tons), later converted into Tons in power BI

     ▪ **Yield** (kg/ha) converted it into ton/ha

3. **Merging Datasets**:

   o Combined multiple crop-specific datasets into a unified table with:

     ▪ Crop

     ▪ Area

     ▪ Production

     ▪ Crop Category

4. **Categorization**:

   o Grouped crops into **Rice**, **Wheat**, and **Pulses** (Chickpea, Pigeonpea, Minor Pulses).

   o All other crops grouped as **"Other"**.

5. **Data Cleaning**:

   o Removed records where **Area**, **Production**, or **Yield** values were zero.

   o Ensured consistent naming conventions and formatting.

6. **Final Output**:

   o Produced a cleaned dataset df_merged.

   o Exported as **Cleaned_Crops_data.csv** in Crop Yield prediction Code file to be used in dashboard

**Insights** *(from the Power BI Report)*

- **Top Contributing Crops**: Rice and Wheat dominate total production across most states.

- **Seasonality Patterns**: Kharif season shows higher overall production volumes compared to Rabi and other seasons.

- **State-wise Yield Leaders**: Certain states consistently outperform in yield, especially for Rice in Kharif and Wheat in Rabi.

- **Pulses Performance**: Pulses contribute less to total production but show high yield variability, indicating potential for improvement.

- **Area vs Yield Trade-offs**: Larger cultivation area doesn't always translate to higher yield, suggesting efficiency gaps.