# Crop Yield Prediction & Optimization

M. Ramya.

sahasrapramod@gmail.com.

GitHub link: Ramyamarripedda

## 1) Project overview:

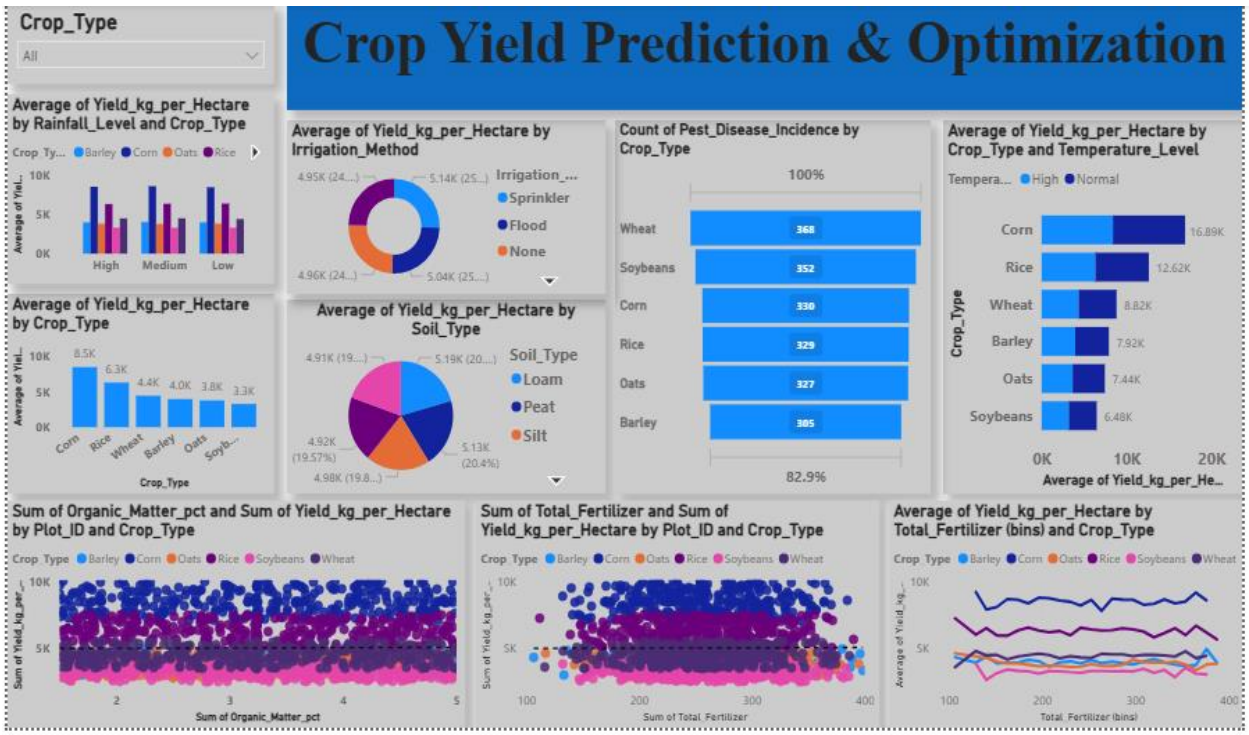
"This project visualizes crop yield trends across different crop types, soil conditions, NPK inputs, and environmental factors using Power BI."

## 2) Dataset Description:

### i) List all columns:

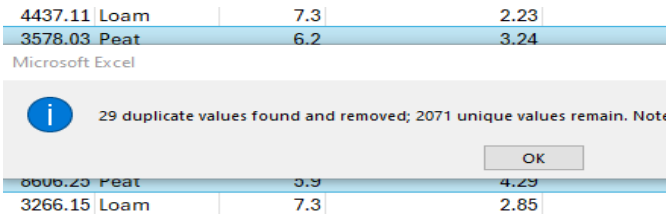
Plot_ID, Year, Crop_Type, Yield_kg_per_Hectare, Soil_Type, Soil_pH, Organic_Matter_pct, Nitrogen_kg_per_Hectare, Phosphorus_kg_per_Hectare, Potassium_kg_per_Hectare, Avg_Rainfall_mm, Avg_Temperature_C, Pest_Disease_Incidence, Irrigation_Method.

## 3) Dashboard Visualizations:

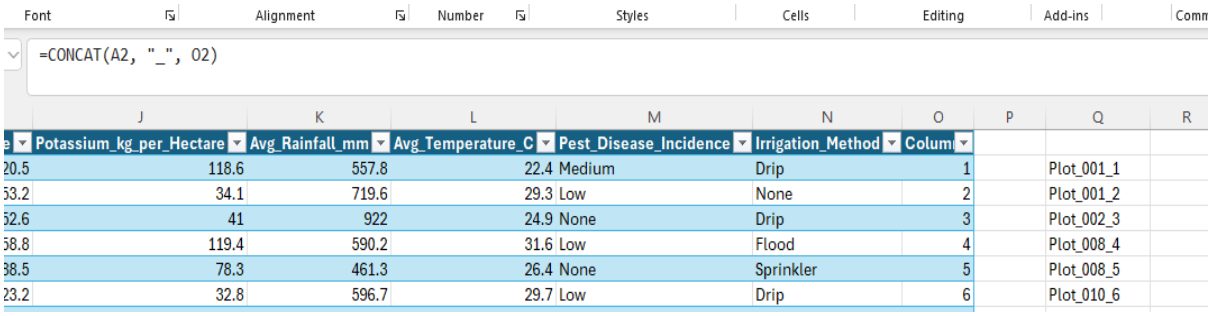



## 4) Data Preparation: (Cleaning, transformation)

### i) Deleted duplicate values:

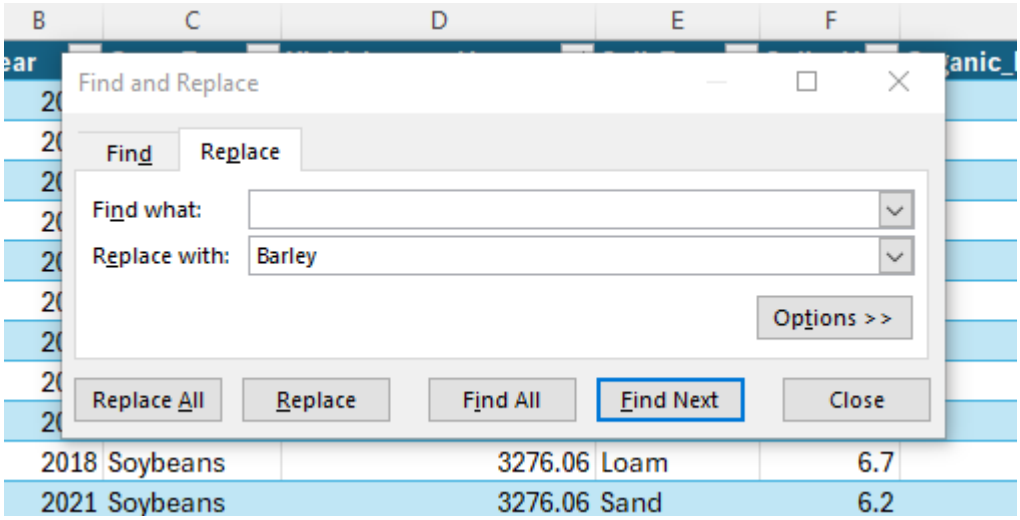


**ii) (Plot_ID):** A new column was created with sequential values (1, 2, 3, …) and concatenated with the Plot_ID to uniquely identify and eliminate confusion caused by duplicate Plot_ID entries:




| | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|
| | Potassium_kg_per_Hectare | Avg_Rainfall_mm | Avg_Temperature_C | Pest_Disease_Incidence | Irrigation_Method | Colum | | | |
| 20.5 | 118.6 | 557.8 | 22.4 | Medium | Drip | 1 | | Plot_001_1 | |
| 53.2 | 34.1 | 719.6 | 29.3 | Low | None | 2 | | Plot_001_2 | |
| 52.6 | 41 | 922 | 24.9 | None | Drip | 3 | | Plot_002_3 | |
| 58.8 | 119.4 | 590.2 | 31.6 | Low | Flood | 4 | | Plot_008_4 | |
| 38.5 | 78.3 | 461.3 | 26.4 | None | Sprinkler | 5 | | Plot_008_5 | |
| 23.2 | 32.8 | 596.7 | 29.7 | Low | Drip | 6 | | Plot_010_6 | |

**iii) (Crop_Type):** The number of records for each crop type was counted, and the average count was calculated. 'Barley', being closest to the average, was used to fill the blank values in the Crop_Type column:

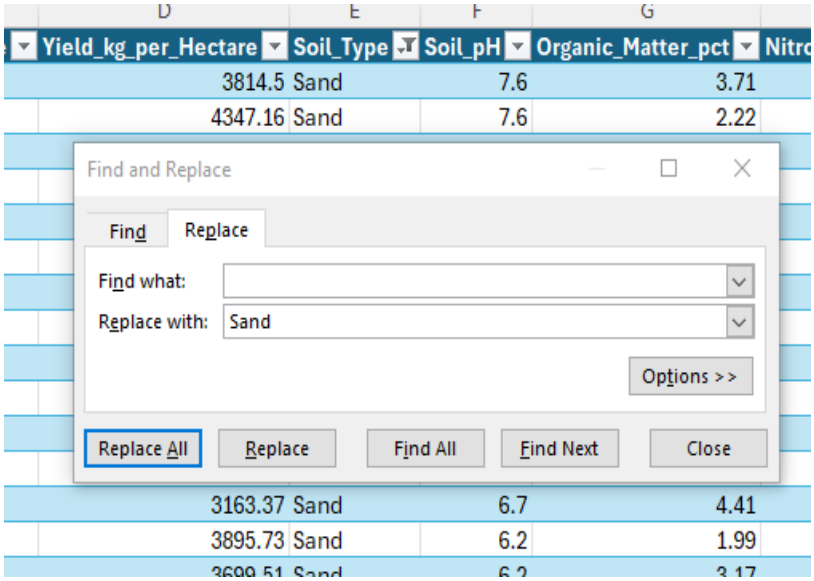



**iv) (Yield_kg_per_Hectare):** Blank values in the Yield_kg_per_Hectare column were filled with the average yield of their respective crop type.

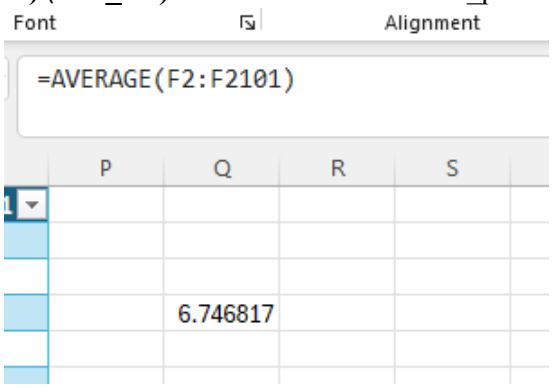
| (Crop_Type) | (Mean_Yield) |
|---|---|
| Barley | 3959.7 |
| Corn | 8481.3 |
| Oats | 3777.3 |
| Rice | 6290.4 |
| Soybeans | 3276.06 |
| Wheat | 4437.11 |

**v) (Soil_type):** The count of each Soil_Type was calculated, and the average was determined. 'Sand', being closest to the average, was used to fill the blank cells in the Soil_Type column.

| (soil_type) | (count) |
|---|---|
| clay | 373 |
| loam | 378 |
| peat | 433 |
| sand | 399 |
| silt | 412 |




**vi) (Soil_Ph):** Blank values in the Soil_pH column were filled with the average value of 6.7.




**vii) (Organic_Matter_pct):** Blank values in the Organic_Matter_pct column were filled with the average value of 3.24.

**viii) (Nitrogen_kg_per_Hectare):** Blank values in the Nitrogen_kg_per_Hectare column were filled with the average value of 123.7.

**ix) (Phosphorus_kg_per_Hectare):** Blank values in the Phosphorus_kg_per_Hectare column were filled with the average value of 58.9.

**x) (Potassium_kg_per_Hectare):** Blank values in the Potassium_kg_per_Hectare column were filled with the average value of 75.4.

**xi) (Avg_Rainfall_mm):** Blank values in the Avg_Rainfall_mm column were filled with the average value of 905.2.

**xii) (Avg_Temperature_C):** Blank values in the Avg_Temperature_C column were filled with the average value of 24.9.

## 5) Questions:

1)What is the average yield_kg_per_hectar by crop type?

2) Which siol_type produces the highest average yield?

3) Does organic_matter_pct positively correlate with yield?

4) What is the average yield per crop under different rainfall levels?

5) Which crop_type perform best in high_temperature conditions?

6) Which crop_type has the highest pest/disease incidence on average?

7) How does yield_kg_per_hectar vary by irrigation_method?

8) How does the combination NPK input (nitrogen, phosphorus, and potassium combined) affect crop yield across different crop_type?

## 6) DAX Functions:

1)Temperature_Level = IF('Crop'[Avg_Temperature_C]> 30, "High", "Normal")

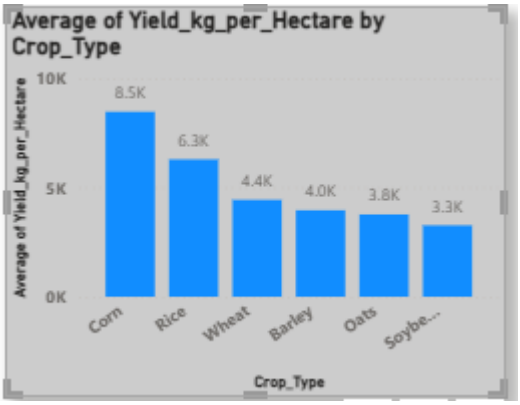
2)Total_Fertilizer = 'Crop'[Nitrogen_kg_per_Hectare] +'Crop'[Phosphorus_kg_per_Hectare] + 'Crop'[Potassium_kg_per_Hectare]

3) Rainfall_Level = SWITCH (TRUE (), 'Crop'[Avg_Rainfall_mm] < 500, "Low",
'Crop'[Avg_Rainfall_mm] >= 500 && 'Crop'[Avg_Rainfall_mm] < 1000, "Medium",
'Crop'[Avg_Rainfall_mm] >= 1000, "High","Unknown")

These are the DAX functions used in this project

## 7) Visualizations:

**i) Average Yield by Crop Type:** A **Clustered column chart** showing the average yield for each Crop_Type. This helps identify which crops perform best overall. (Corn & Rice Crop Type perform best overall).

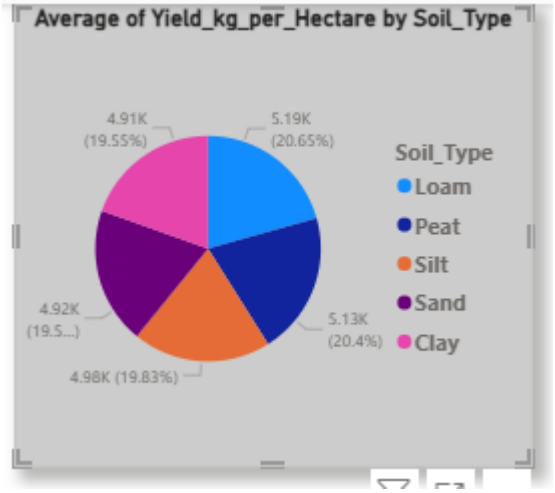



**x-axis = crop_type**

**y-axis = yield_kg_per_hectare**

**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

**ii) Yield by Soil Type:** A **Pie chart** comparing average yield across different Soil_Type values to determine the most productive soil. (Lome & Peat are the most productive soils).

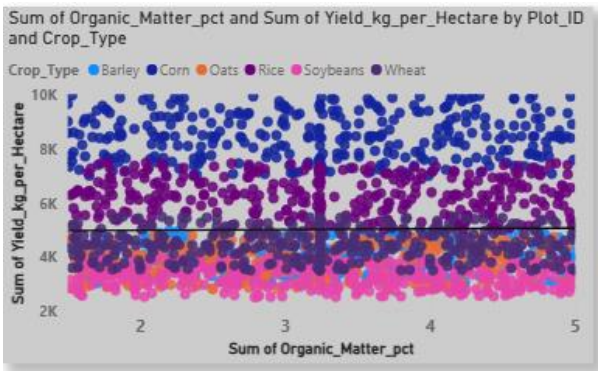



**legend = soil_type**

**values = yield_kg_per_hectare**

**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

**iii) Organic Matter vs Yield:** A **scatter plot** with a trendline, used to observe correlation between Organic_Matter_pct and Yield_kg_per_Hectare.




**x-axis = Organic_matter_pct**

**y-axis = yield_kg_per_hectare**

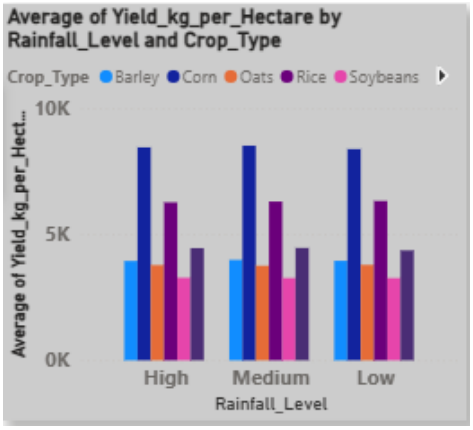
**Add a trendline (click on the scatter chart, go to analytics pane, add a trend line)**

a) If the trend line goes **upward**, it indicates a **positive correlation.**

b) If the trend line goes **downward**, it shows a **negative correlation**.

c) A **flat** trend line indicates no significant relationship.

**iv) Yield under Rainfall Levels:** A **clustered column chart** showing how yield varies across rainfall categories for each crop type.( Crops grown under moderate rainfall conditions yielded better on average than those under low or high rainfall)




**x-axis = Rainfall_Level**

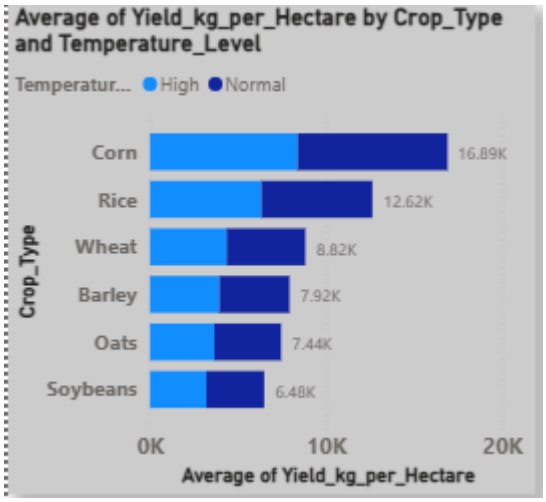
**y-axis = yield_kg_per_hectare**

**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

a) The Avg_Rainfall_mm column was categorized into ranges by creating a new calculated column, which groups rainfall levels into categories such as Low, Moderate, High, by using DAX function.

```
Rainfall_Level = SWITCH (TRUE (), 'Crop'[Avg_Rainfall_mm] < 500, "Low",
'Crop'[Avg_Rainfall_mm] >= 500 && 'Crop'[Avg_Rainfall_mm] < 1000, "Medium",
'Crop'[Avg_Rainfall_mm] >= 1000, "High","Unknown")
```

**v) Yield by Temperature Conditions:** The stacked bars allow comparison of yield contributions from each crop type across temperature ranges. (Rice and Corn, performed better under high-temperature conditions, while others yielded more under normal temperatures)




**y-axis = Crop_type.**
**x-axis = yield_kg_per_hectare.**
**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

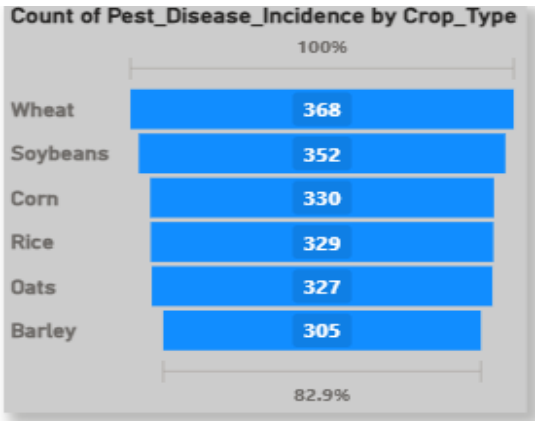
**Legend = temperature_value.**

**a)** The Avg_Temperature_C column was categorized into temperature levels by creating a new calculated column. Values were grouped into categories such as 'Normal' and 'High' based on defined thresholds, by using DAX function.

```
Temperature_Level = IF('Crop'[Avg_Temperature_C]> 30, "High", "Normal")
```

**vi) Pest Incidence by Crop:** A funnel chart was used to display the average pest/disease incidence for each crop type. The chart ranks crop from highest to lowest based on their vulnerability, helping highlight those that require more pest control attention.

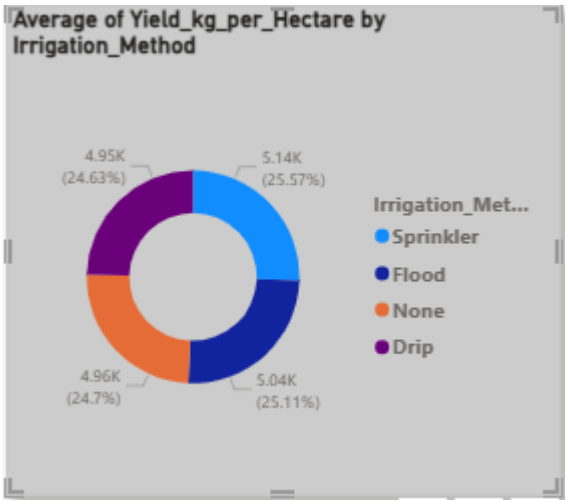
(Wheat and Soybean had the highest average pest/disease incidence)




Category = Crop_Type.

Values = Count of Pest_Disease_Incidence.

**vii) Yield by Irrigation Method:** A donut chart was used to compare the average yield under different irrigation methods. This visualization highlights which irrigation technique results in higher productivity. (Sprinkler irrigation resulted in the highest average yields)




**legend = Irrigation_method.**

**Values = yield_kg_per_hectare.**

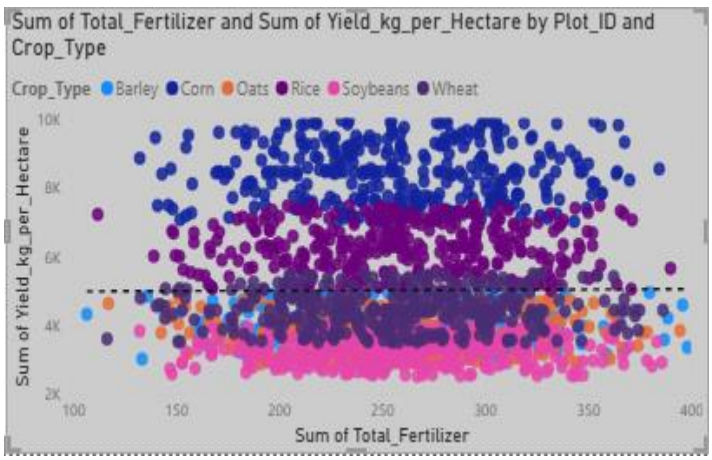
**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

**viii) NPK Input vs Yield by Crop:**

1) A scatter chart was used to visualize the relationship between total fertilizer input and yield, helping identify patterns or thresholds where fertilizer application begins to have diminished returns
2) A **line chart** using Total_Fertilizer_Bin on the X-axis to see how fertilizer levels affect yield across crops.
3) A new calculated column Total_Fertilizer was created by summing the values of nitrogen, phosphorus, and potassium inputs per hectare. This column represents the combined NPK fertilizer input used for each crop record and is used to analyse its effect on crop yield, by DAX function.

Total_Fertilizer = 'Crop'[Nitrogen_kg_per_Hectare]
+'Crop'[Phosphorus_kg_per_Hectare] + 'Crop'[Potassium_kg_per_Hectare]

**1) scatter chart**




**values = Plot_ID**

**x-axis = Total_Fertilizer**

**y-axis = yield_kg_per_hectare.**

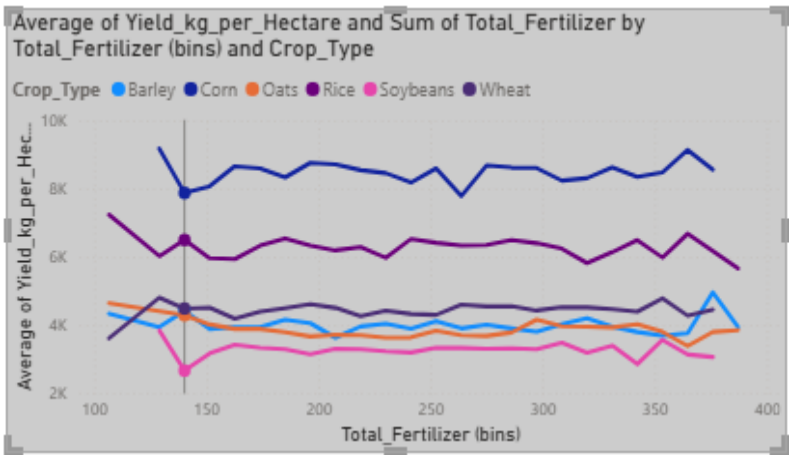
**Legend = Crop_Type.**

a) If the trend line goes **upward**, it indicates a **positive correlation.**

b) If the trend line goes **downward**, it shows a **negative correlation**.

c) A **flat** trend line indicates no significant relationship.

2) **line chart:** The line chart shows how crop yield changes with different fertilizer levels. It helps find the best fertilizer range that gives higher yield for each crop.




**x-axis = Total_Fertilizer.**

**y-axis = yield_kg_per_hectare.**

**(Convert sum of yield_kg_per_hectare to Average of yield_kg_per_hectare)**

**Legend = Crop_Type.**

To create bins for Total_Fertilizer, go to the Visualization pane, drag Total_Fertilizer to the X-axis, then right-click on it and select 'New group'. In the group settings, choose 'Bin Type' as 'Number of bins' and set the desired number.

## 8) Insights:

1. **Crop Type vs Yield**
   Corn and Rice showed the highest average yields, indicating they are the most productive crops in the dataset.
2. **Soil Type Impact**
   Loam and Peat soils were associated with better yields, while sandy soils showed relatively lower performance.
3. **Organic Matter Influence**
   A positive trend was observed between organic matter percentage and crop yield, suggesting that increasing soil organic content can improve productivity.
4. **Rainfall vs Yield**
   Crops grown under moderate rainfall conditions yielded better on average than those under low or high rainfall, indicating an optimal moisture range.
5. **Temperature Conditions**
   Some crops, like Rice and Corn, performed better under high-temperature conditions, while others yielded more under normal temperatures.
6. **Pest and Disease Incidence**
   Wheat and Soybean had the highest average pest/disease incidence, highlighting the need for more protection measures for these crops.

7. **Irrigation Method**
   Sprinkler irrigation resulted in the highest average yields, showing its efficiency over methods like flood or Drip irrigation.
8. **Fertilizer (NPK) Effect**
   A clear positive relationship was seen between total NPK input and yield up to a certain level. Beyond that, yield gains slowed or plateaued, indicating a point of diminishing returns.
9. **Total Fertilizer Bin Analysis**
   Line and column charts grouped by fertilizer bins revealed that most crops achieved optimal yield in the 200–300 kg/ha input range.
10. **Missing Data Handling**
    Data cleaning steps, such as filling missing values based on averages or minimum frequency categories, ensured more accurate and consistent analysis.