***Crop Recommendation System***

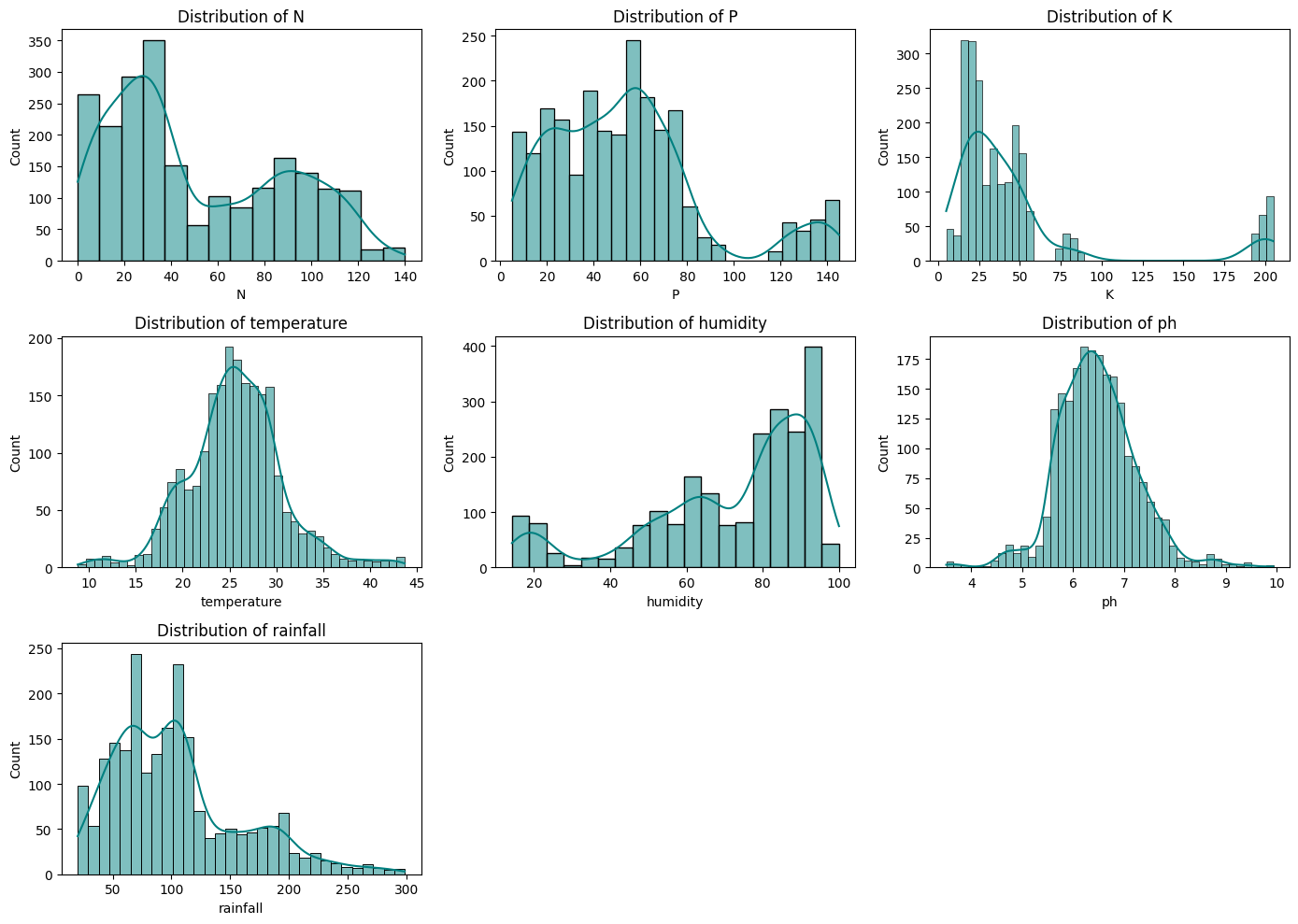
# Exploratory Data Analysis :

## **Count plot** :

From the crop distribution plot, we can see the following insights:

* **Balanced Dataset**: The plot shows that the dataset is remarkably balanced. Each of the 22 unique crop types has exactly 100 data points.
* **Equal Representation**: This equal representation of each crop type is ideal for training a classification model, as it prevents the model from being biased towards any particular crop due to an uneven distribution of data.

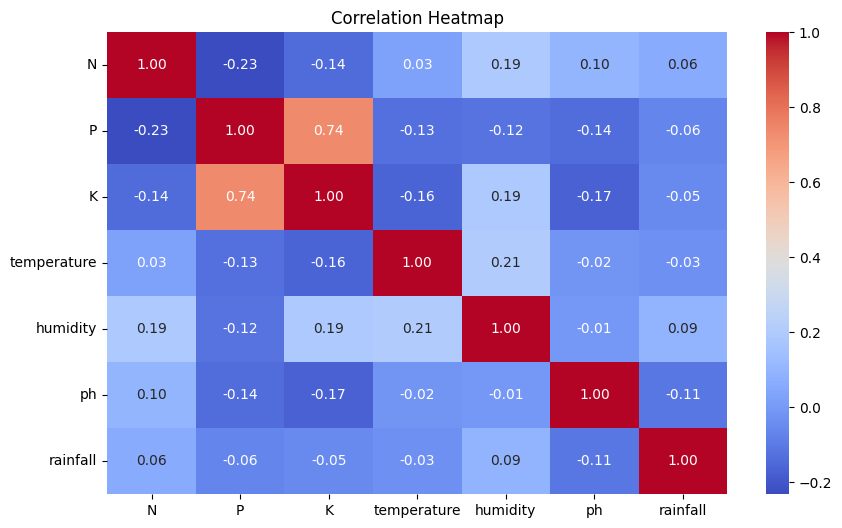
## **Distribution Plot** :



From the distribution plots of the features, we can gain the following insights:

* **N (Nitrogen)**: The distribution of Nitrogen appears to be somewhat multimodal, with peaks at lower and higher values. This suggests that there might be different groups of crops with distinct nitrogen requirements.
* **P (Phosphorus)**: The Phosphorus distribution also shows multiple peaks, indicating varying phosphorus levels across the dataset, likely corresponding to different crop needs.
* **K (Potassium)**: The Potassium distribution is skewed to the right, with a large concentration of values at the lower end and a few crops requiring much higher levels.
* **Temperature**: The temperature distribution seems roughly bell-shaped, indicating a concentration of values around the mean temperature.
* **Humidity**: The humidity distribution is skewed towards higher values, with a significant number of observations having high humidity. There is also a smaller peak at lower humidity levels.
* **pH**: The pH distribution appears to be relatively normal (bell-shaped), centered around a neutral pH, with some spread towards more acidic and alkaline conditions.
* **Rainfall**: The rainfall distribution is skewed to the right, suggesting that while many areas have lower rainfall, some locations experience significantly higher amounts.

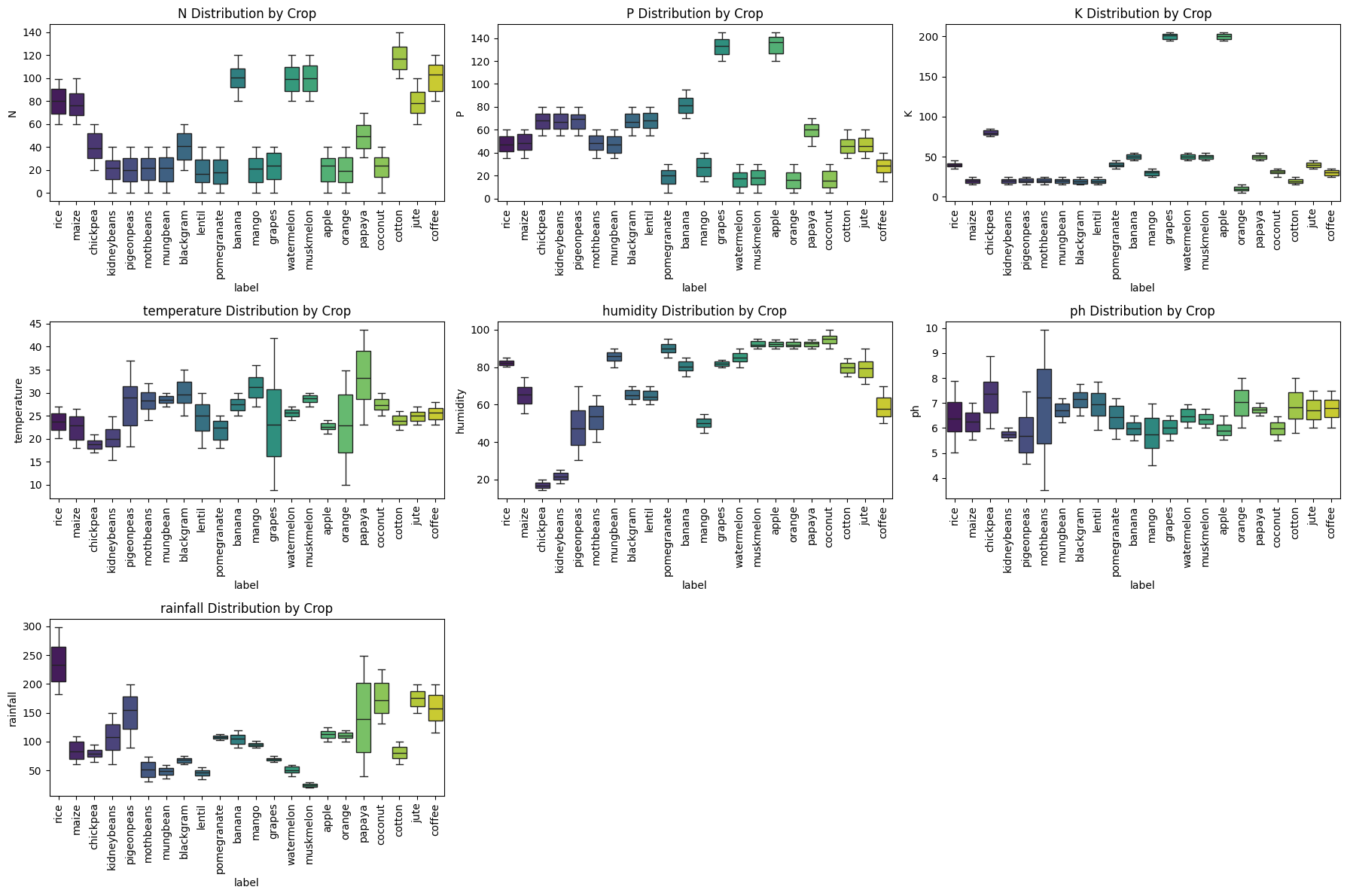
## **Correlation Heatmap** :



Based on the correlation heatmap, insights:

* **Moderate Positive Correlation between P and K**: There's a moderate positive correlation (0.74) between Phosphorus (P) and Potassium (K). This suggests that soils high in phosphorus tend to also be high in potassium, and vice-versa.
* **Moderate Negative Correlation between N and P, and N and K**: There are moderate negative correlations between Nitrogen (N) and Phosphorus (P) (-0.23), and Nitrogen (N) and Potassium (K) (-0.14). This indicates that as nitrogen levels increase, phosphorus and potassium levels tend to slightly decrease.
* **Weak Correlations**: Most other features (temperature, humidity, ph, and rainfall) show relatively weak correlations with each other and with N, P, and K. The correlation coefficients are close to zero, suggesting that these features are largely independent of each other. For example, temperature has a weak positive correlation with humidity (0.21) but very weak correlations with N, P, K, pH, and rainfall.

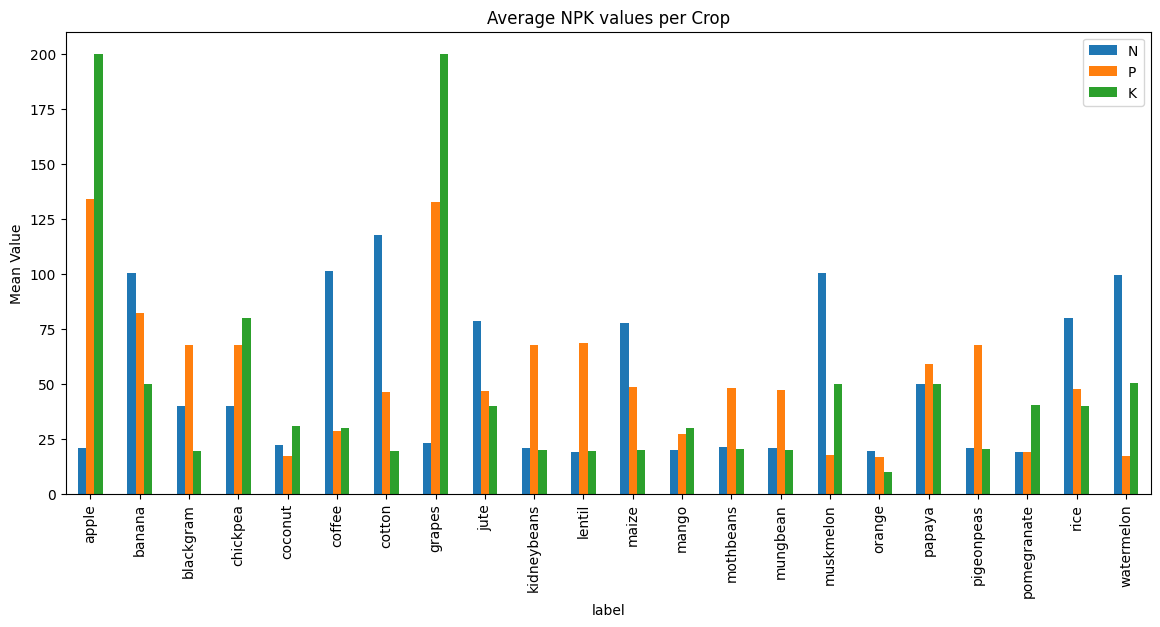
## **Box Plot** :



The box plots show how the distribution of each nutrient (N, P, K), temperature, humidity, pH, and rainfall varies for each crop type. We can extract several insights from these visualizations:

* **Nutrients (N, P, K)**: Different crops have distinct requirements for Nitrogen (N), Phosphorus (P), and Potassium (K). For example, rice and cotton seem to require higher levels of N, while certain fruits like grapes and mangoes might thrive with different nutrient balances. Potassium levels show significant variation across crops, with some like banana and grapes showing higher median K values.
* **Temperature and Humidity**: These plots indicate the optimal temperature and humidity ranges for each crop. For instance, coffee appears to prefer a narrower temperature range compared to other crops. Humidity levels are generally high for many crops, but some like chickpea and kidneybeans seem to tolerate lower humidity.
* **pH**: The box plot for pH shows the preferred soil pH range for each crop. Most crops appear to grow in a slightly acidic to neutral pH range (around 6 to 7), but some like citrus fruits might tolerate more acidic conditions.
* **Rainfall**: The rainfall plot highlights the water requirements of different crops. Rice, for example, requires significantly higher rainfall compared to crops like chickpea or kidneybeans.

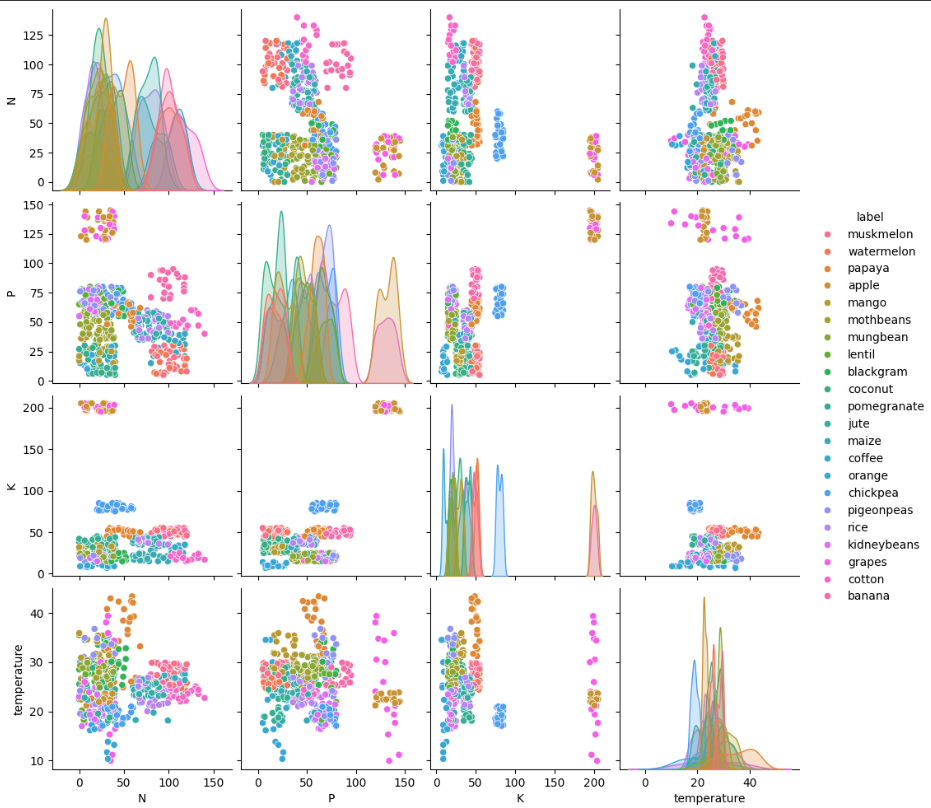
## **Bar Plot** :



The bar plot showing the average NPK values per crop provides clear insights into the nutrient needs of different crops. Here are some observations:

* **High Potassium (K) for Apple and Grapes**: Apple and grapes stand out with significantly higher average potassium levels compared to other crops. This suggests these fruits require a substantial amount of potassium for optimal growth.
* **High Nitrogen (N) for Certain Crops**: Crops like cotton, coffee, muskmelon, banana, and rice show relatively high average nitrogen levels. This indicates that nitrogen is a crucial nutrient for these crops.
* **Varied Phosphorus (P) Requirements**: Phosphorus levels vary across crops. Chickpea, kidneybeans, lentil, and pigeonpeas have relatively similar and moderate average phosphorus levels. Apple and grapes also show high phosphorus levels, similar to their potassium needs.
* **Low Nutrient Requirements for Some Crops**: Some crops, like orange and mango, appear to have lower average requirements for all three nutrients (N, P, and K) compared to others.
* **Distinct Nutrient Profiles**: Each crop generally has a distinct combination of average N, P, and K values, highlighting their specific nutritional needs. This differentiation is key for recommending the right crop based on soil nutrient levels.

## **Pair Plot** :



The pair plot provides a visual exploration of the relationships between pairs of features (N, P, K, and temperature) and how these relationships vary across different crop types. Here are some insights from the pair plot:

* **Crop Clustering in Feature Space**: We can observe that different crop types tend to form distinct clusters or groups in the scatter plots. This indicates that combinations of these features (N, P, K, and temperature) can help differentiate between different crops. For example, in the N vs. K scatter plot, several crops occupy clearly separated regions.
* **Separation based on N, P, and K**: The scatter plots involving N, P, and K show good separation for many crop types. This reinforces the importance of nutrient levels in determining suitable crops. Crops with similar nutrient requirements tend to cluster together.
* **Temperature's Role in Differentiation**: While temperature shows some variation across crops (as seen in the diagonal histograms and scatter plots involving temperature), its ability to separate crops appears less pronounced compared to the nutrient features (N, P, K). Some crops might have overlapping temperature ranges.
* **Relationships between Features**: The scatter plots also show the relationships between pairs of features. For instance, the P vs. K scatter plot shows a somewhat positive relationship, consistent with the correlation heatmap, and we can see how different crops are positioned along this relationship.
* **Potential for Classification**: The visible clustering of different crop types in the feature space suggests that these features are good predictors for crop recommendation and that a classification model could be built to predict the crop based on these environmental and nutrient parameters.

# Research or Innovation Idea :

## **“ Adaptive Multi-Crop Planning System “**

### **a. What problem or opportunity does the insight suggest?**

From the dataset analysis, we see that each crop has unique nutrient (N, P, K), rainfall, and pH requirements. Farmers usually grow a **single crop per season**, but continuous cultivation of the same crop or random choices leads to **soil nutrient depletion, reduced yield, and unsustainable farming practices**.

**Opportunity:** Instead of recommending just one suitable crop, there is potential to recommend **multi-crop combinations and crop rotation schedules** that balance soil nutrients, maintain fertility, and improve long-term yield and farmer income.

### **b. How could an R&D team develop a solution?**

1. **Dataset Extension**: Augment the current dataset with crop rotation knowledge (nutrient replenishing crops vs nutrient-consuming crops).
2. **Clustering & Optimization Models**:  
   * Group crops by nutrient usage profiles (e.g., nitrogen-hungry vs nitrogen-restoring).
   * Use optimization algorithms to recommend crop sequences that balance soil health.
3. **Decision Support Dashboard**:  
   * Build a farmer-friendly tool that simulates **multi-year crop planning**.
   * Provide recommendations for intercropping (multiple crops in one season) and rotation (across seasons).
4. **Integration**: Include climate and rainfall predictions so that crop cycles are also weather-resilient.

### **c. Brief steps to validate or test the idea (Optional)**

1. **Pilot Study**: Test crop rotation plans with a small group of farmers.
2. **Compare Results**: Measure soil fertility, crop yield, and profitability against single-crop farming.
3. **Feedback Loop**: Collect farmer feedback and refine the model.
4. **Scaling**: Expand to diverse regions with different soil and climate conditions to improve generalization.

# Comparative Study of Findings :

### **a. Key Patterns or Trends Discovered**

1. **Balanced Dataset** – Each of the 22 crop types has equal representation, avoiding bias in analysis and model training.
2. **Nutrient Requirements Differ** – Some crops (like rice, cotton, and banana) demand high nitrogen, while fruits such as grapes and apples require higher potassium and phosphorus.
3. **Climate Influence** – Rice needs significantly higher rainfall, while crops like chickpea and kidney beans thrive in drier conditions. Coffee and certain fruits are sensitive to narrow temperature or pH ranges.
4. **Distinct Clusters** – Scatter plots and pair plots show that crops form natural clusters based on N, P, K, and climatic conditions, indicating strong separability for classification models.

### **b. Insights Gained from the Dataset**

1. **Soil–Crop Match is Critical** – Crops have unique nutrient and climate “fingerprints” that strongly determine their suitability.
2. **Sustainability Risk** – Continuous cultivation of nutrient-heavy crops can deplete soil fertility if not rotated with nutrient-restoring crops.
3. **Precision Potential** – The dataset provides a strong base for building ML-driven recommendation systems that can suggest crops tailored to soil and weather conditions.
4. **Correlation Clues** – Moderate correlation between phosphorus and potassium, and negative relation with nitrogen, highlights natural trade-offs in soil nutrient balances.

### **c. Rationale Behind the Proposed Idea (Adaptive Multi-Crop Planning System)**

While most existing systems focus on recommending a **single best crop**, the findings reveal that crops differ significantly in nutrient demands and climate needs. This suggests that **long-term planning through crop rotation or intercropping** can:

* Preserve soil fertility by balancing nutrient usage.
* Increase overall yield and reduce dependency on chemical fertilizers.
* Provide farmers with sustainable, climate-resilient cropping strategies.

# Conclusion :

### ***Adaptive Multi-Crop Planning System*** *directly addresses the limitations of single-crop recommendations by leveraging these dataset insights to design* ***nutrient-balanced, profit-maximizing, and eco-friendly crop cycles****.*