Exploratory Data Analysis on Crop Recommendation Dataset

By: Arpita Mishra

Objective:

To understand how soil nutrients and environmental factors influence crop suitability, and extract actionable insights that can help improve agricultural decision-making.

Focus:

- Nutrient analysis
- Climate patterns
- Soil chemistry
- Crop grouping & trends

Dataset Overview

N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.879744	82.002744	6.502985	202.935536	rice
85	58	41	21.770462	80.319644	7.038096	226.655537	rice
60	55	44	23.004459	82.320763	7.840207	263.964248	rice
74	35	40	26.491096	80.158363	6.980401	242.864034	rice
78	42	42	20.130175	81.604873	7.628473	262.717340	rice

The dataset contains 2200 records of different crops and their environmental requirements.

Includes 22 unique crops such as rice, maize, cotton, chickpea, kidney beans, grapes, apple, banana, etc.

Each record provides soil and climate conditions required by a crop.

7 key features are included:

- N (Nitrogen)
- P (Phosphorus)
- K (Potassium)
- Temperature (°C)
- Humidity (%)
- pH
- Rainfall (mm)

Target variable: label (Crop)

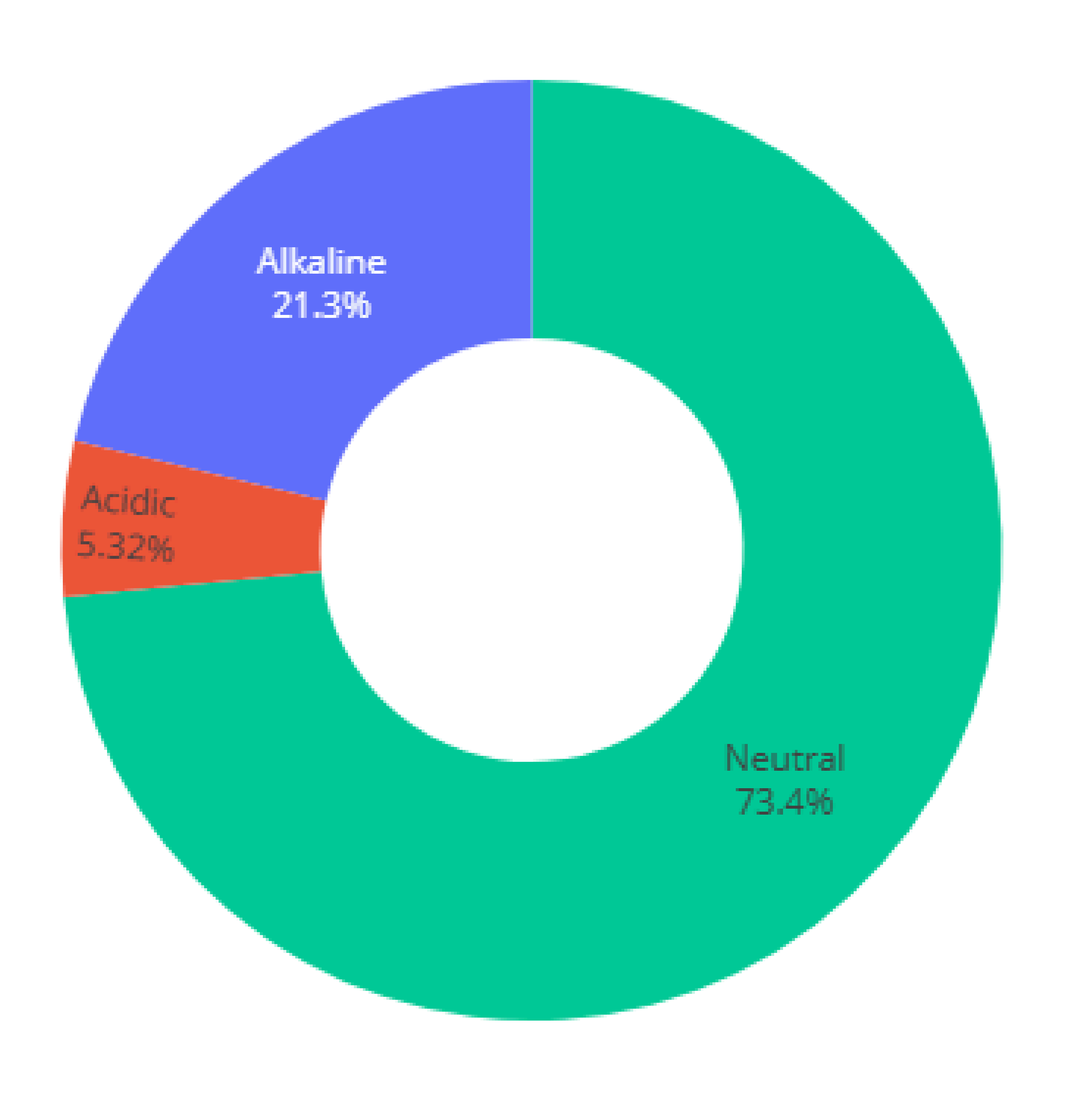
Feature Engineering

1. Nutrient Balance Score
Defined as the average of N, P, and K.
Helps understand whether a crop requires a balanced nutrient profile or is biased toward specific nutrients.

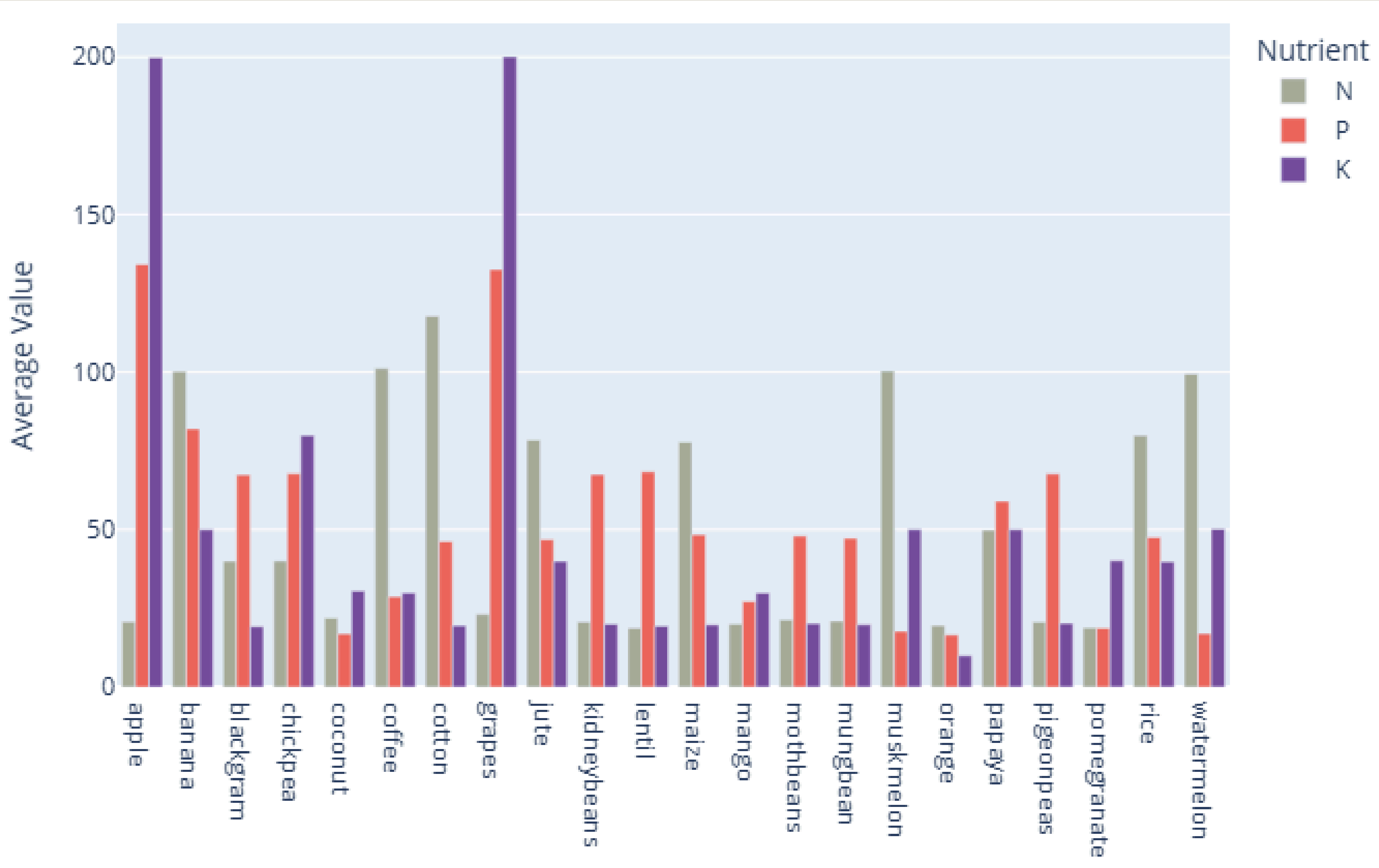
2. Moisture Index
Calculated as: Humidity × Rainfall
Represents the overall moisture availability for crops, combining atmospheric humidity and rainfall.

3. Soil pH Category
Converted numeric pH into three categories:
- Acidic (pH < 6.5)
- Neutral (6.5 ≤ pH ≤ 7.5)
- Alkaline (pH > 7.5)

Useful for grouping crops by soil compatibility.



How do the nutrient patterns (N, P, K) differ from crop to crop?

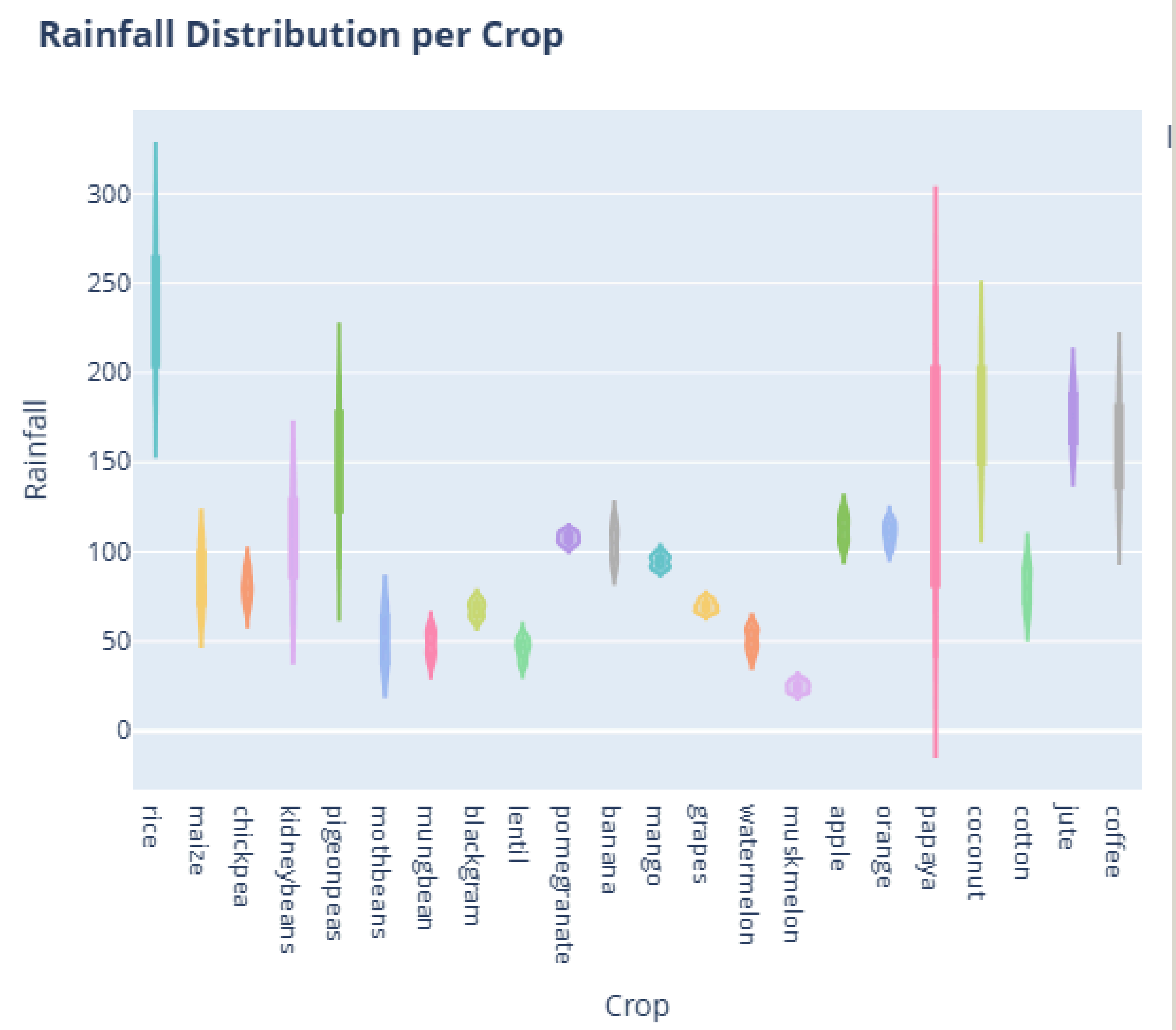


Some crops like grapes and apple show higher Potassium(K) levels indicating they rely more on K for fruit development.

Crops like watermelon, muskmelon, cotton, coffee, maize, and rice show higher Nitrogen (N) levels. This indicates they require stronger vegetative growth and benefit from nitrogen-rich soil.

Phosphorus (P) levels vary moderately, helping identify crops that need stronger root systems or flowering support.

How does rainfall preference vary across crops?



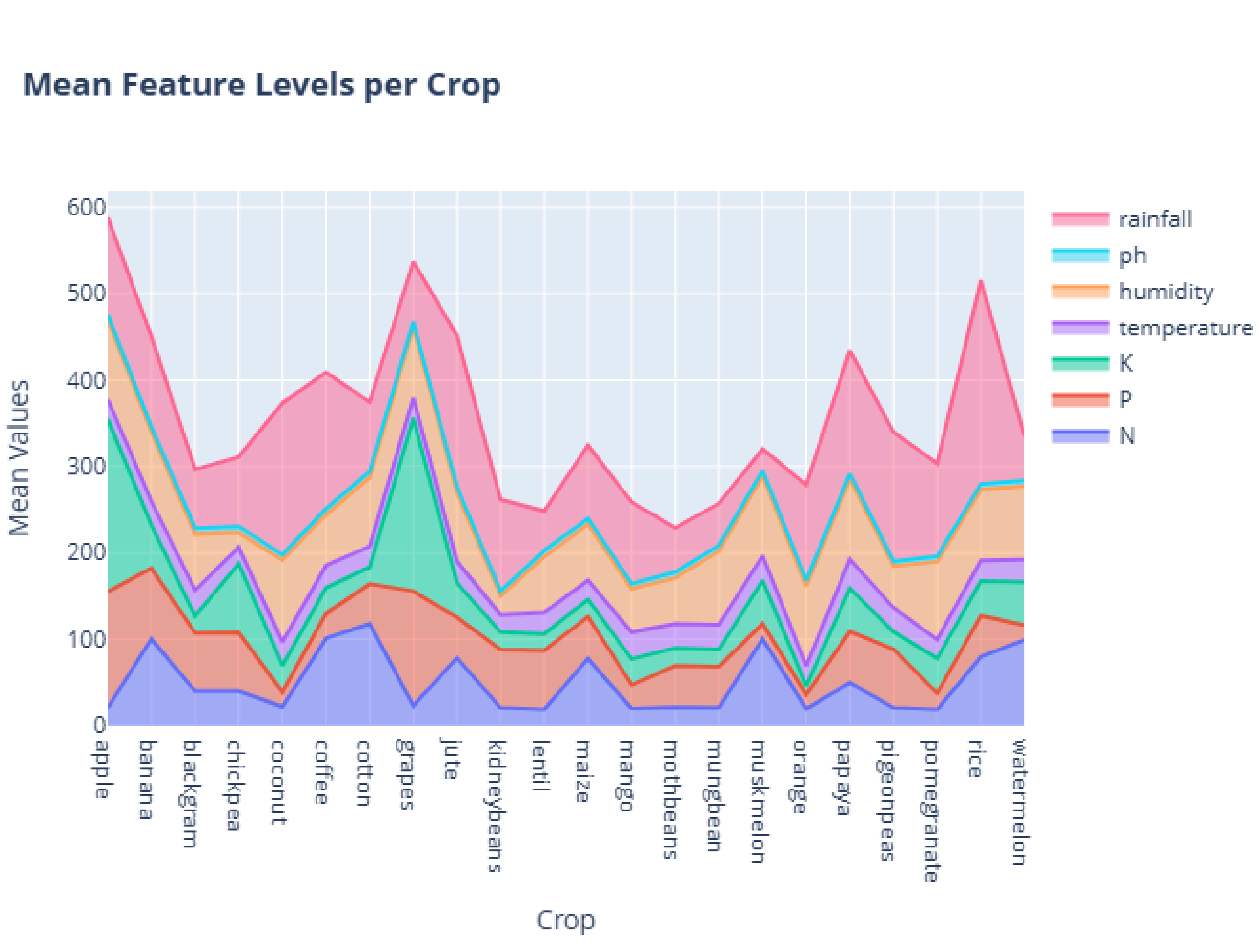
The violin plot shows the spread and density of rainfall values for each crop, helping identify which crops thrive in wetter vs drier environments.

Some crops like rice and coconut show consistently high rainfall ranges, indicating they prefer more moisture-rich climates.

Others display lower or narrower rainfall distributions, suggesting they grow well in moderate or dry regions.

Papaya has a wide rainfall distribution, covering both low and high rainfall values.
This indicates papaya is a versatile crop with high adaptability to different moisture levels.

How do overall environmental factors differ across crops?



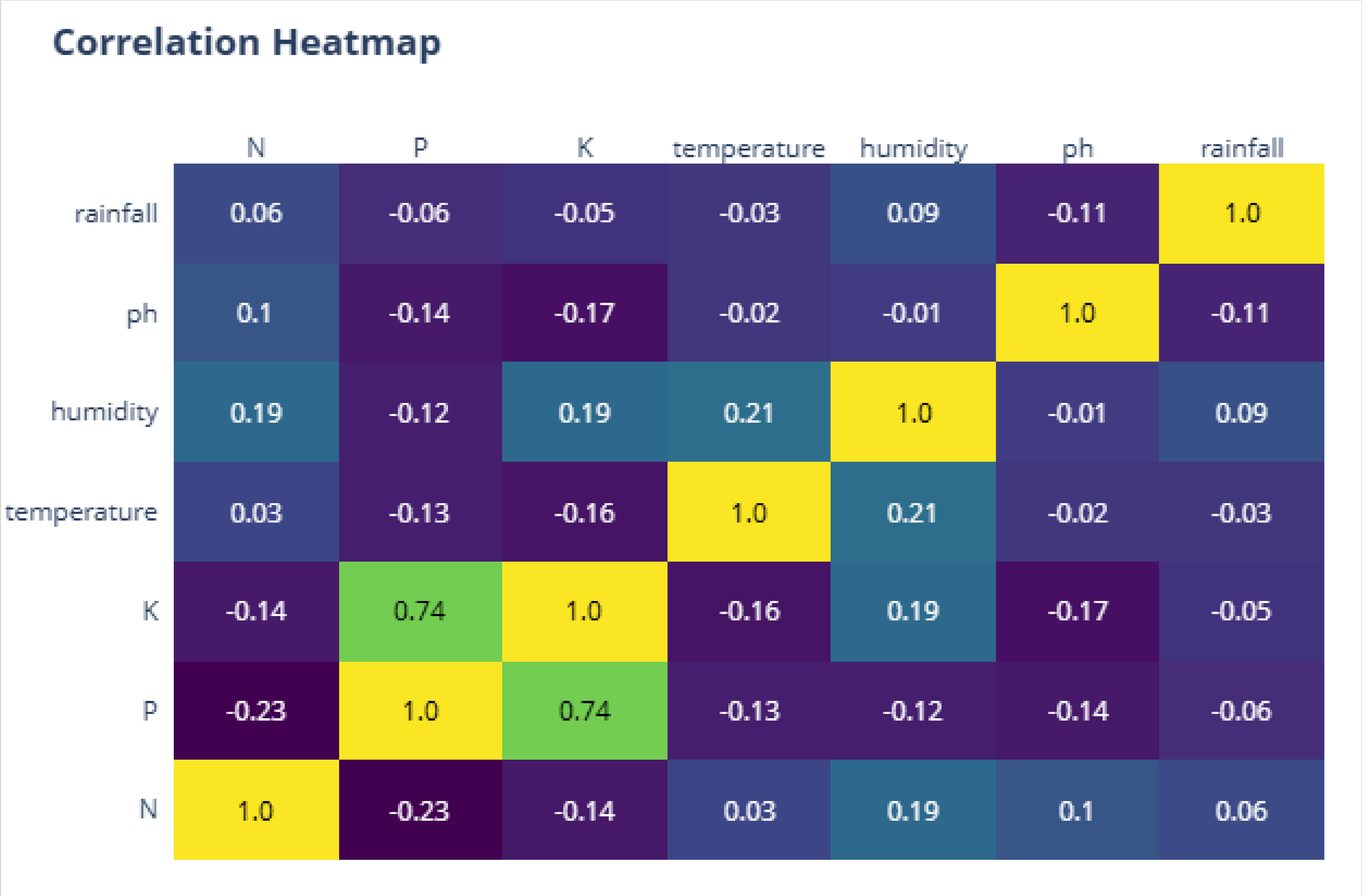
The stacked area chart gives a holistic view of each crop by combining all key features (N, P, K, temperature, humidity, pH, rainfall).

Crops with larger stacked areas require higher overall environmental support (nutrient-rich soil, moisture, suitable temperature).

Crops with smaller areas appear less demanding, meaning they can grow with simpler or moderate conditions.

Differences in layer heights help identify whether a crop is more influenced by nutrients, climate, or soil factors.

Relationships Between Environmental & Nutrient Features

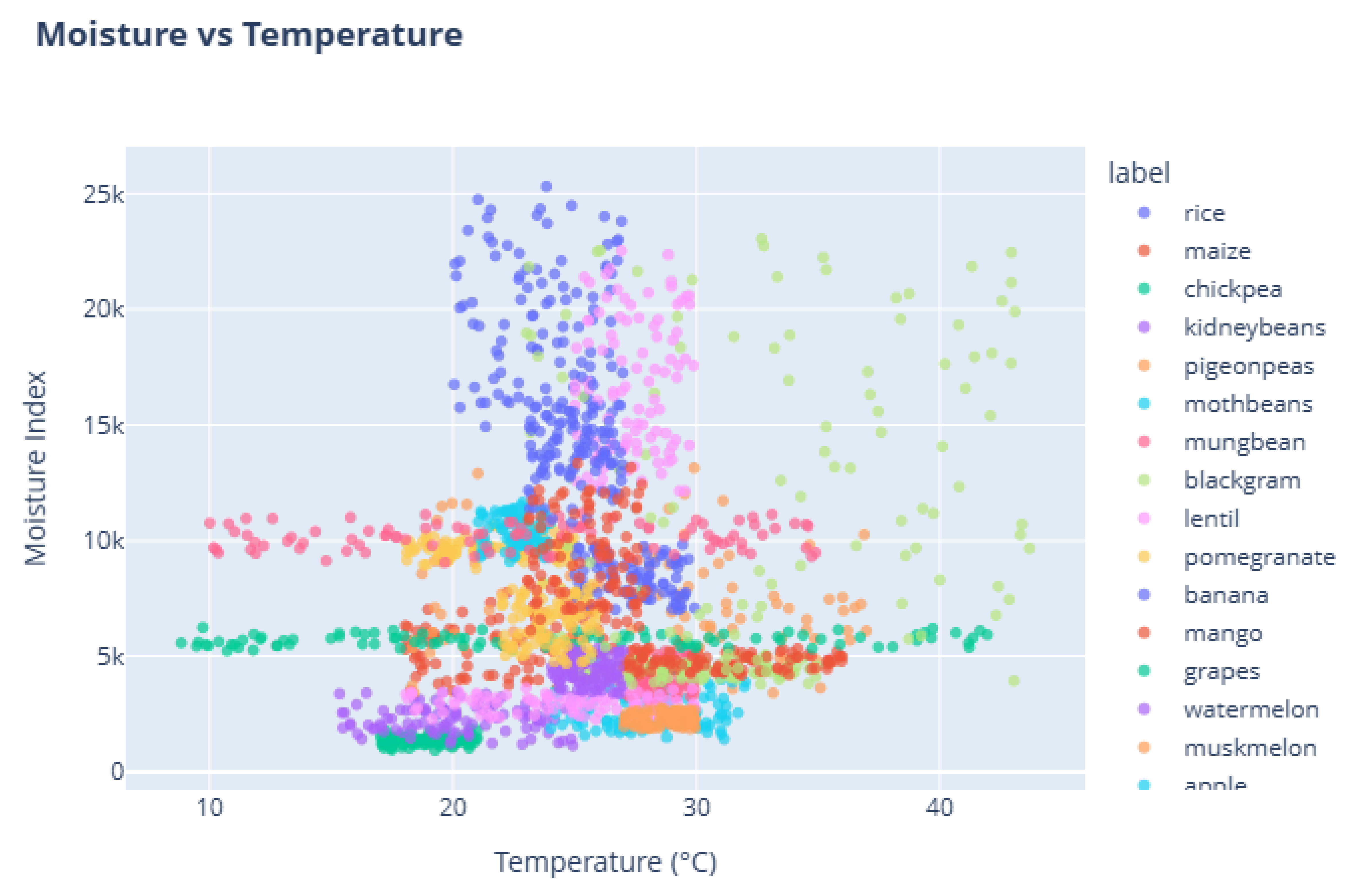


The heatmap highlights how different features move together helping identify patterns, dependencies, and possible interactions.

Strong positive correlations suggest features that increase together, while negative correlations indicate opposite trends.

This helps understand whether nutrient levels (N, P, K) influence each other or if temperature, humidity, and rainfall follow similar climate patterns.

Are there natural groups based on moisture and temperature?



This plot reveals natural groupings of crops - some fall into warm–humid zones, others into hot–dry or moderate–moist regions.

Crops lying close together indicate similar climatic requirements, making them part of the same cluster.

Crops far apart show distinct environmental needs, meaning they thrive under different growing conditions.

This visualization is insightful for multi-crop planning and regional cultivation decisions.

Key Takeaways / Conclusion

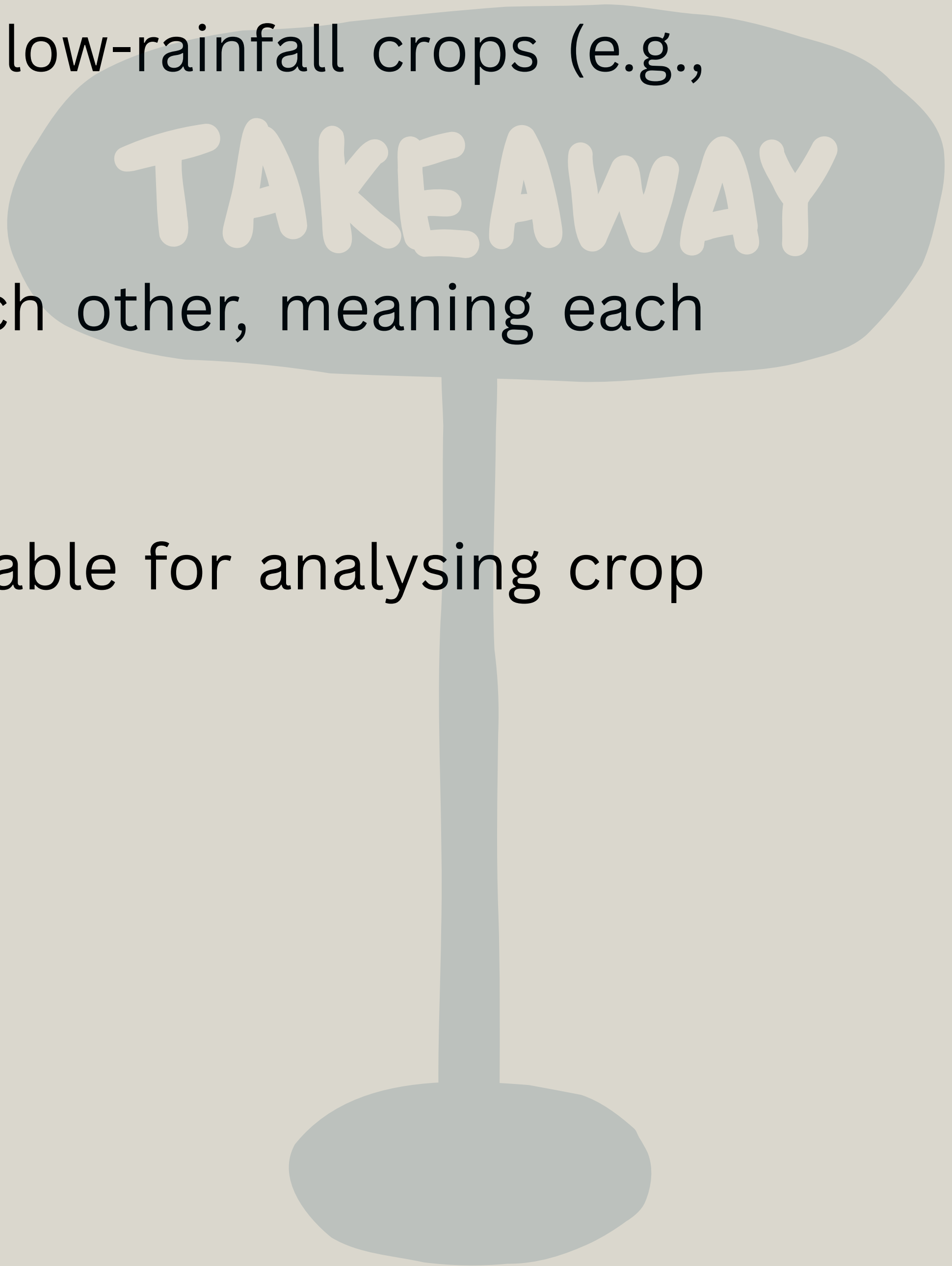
The dataset shows clear variation in soil nutrients (N, P, K) meaning different crops prefer different nutrient levels, which helps in strong feature separation.

Temperature and Humidity play an important role in crop suitability, with warm–humid conditions favoring crops like rice and papaya.

Rainfall acts as a strong deciding factor - high-rainfall crops (e.g., rice) and low-rainfall crops (e.g., wheat, lentil) form clearly distinct groups.

Correlation analysis shows that N, P, K are not strongly correlated with each other, meaning each nutrient contributes independently to crops.

The dataset shows well-defined patterns across crops, making it highly suitable for analysing crop patterns for better decision making.



Thank You.