

# Recomendación Inteligente de Cultivos

Aplicando Ciencia de Datos al Sector Agrícola

Autor: Juan Pablo Arbelaez Otalvaro  
Curso de  
Machine Learning – Talento Tech

# Introducción La Agricultura en la Era Digital

## Contexto Actual

La agricultura moderna se enfrenta a desafíos críticos como el cambio climático, la escasez de recursos hídricos y la necesidad apremiante de aumentar la productividad para alimentar a una población creciente. Estos factores demandan soluciones innovadoras y eficientes.

## Nuestro Propósito

Este proyecto propone la aplicación de técnicas avanzadas de Machine Learning para desarrollar un sistema de recomendación que sugiera el cultivo más idóneo. El objetivo es optimizar las decisiones agrícolas basándose en las condiciones específicas del terreno y el clima, maximizando así el rendimiento y la sostenibilidad.



# Impacto del Proyecto: Agricultura Inteligente y Sostenible



## Toma de Decisiones Basada en Datos

Empodera a los agricultores con información precisa para optimizar la selección de cultivos, minimizando riesgos y maximizando la rentabilidad.



## Optimización de Recursos

Permite un uso más eficiente de insumos cruciales como fertilizantes y agua, contribuyendo a la sostenibilidad ambiental y reduciendo costos operativos.



## Fomento de la Sostenibilidad

Promueve prácticas agrícolas más inteligentes y resilientes, adaptándose a las condiciones cambiantes del entorno y asegurando la viabilidad a largo plazo del sector.

# Descripción del Dataset: La Base de Nuestro Modelo

## Estructura del Dataset

Se utilizó un dataset completo con 2200 registros, cada uno representando un conjunto de condiciones ambientales y del suelo. El dataset está compuesto por 8 columnas, que capturan datos reales y relevantes para la agricultura.

## Características Detalladas

- **N, P, K:** Niveles de nutrientes (Nitrógeno, Fósforo, Potasio) (mg/kg).
- **temperature:** Temperatura ambiental en grados Celsius (°C).
- **humidity:** Humedad relativa en porcentaje (%).
- **ph:** Nivel de acidez o alcalinidad del suelo.
- **lluvia:** Cantidad de lluvia registrada en milímetros (mm).
- **Cultivo:** La variable objetivo, indicando el cultivo ideal para las condiciones dadas.



# Objetivo del Modelo: Clasificación Supervisada de Cultivos



## Cultivos

Nuestro modelo está diseñado para clasificar entre más de 20 opciones de cultivos diferentes, que incluyen variedades de Manzana, banano, frijol negro, garbanzo, coco, café, algodón, uvas, yute, frijol rojo (o frijol riñón), lenteja, maíz, mango, frijol polilla, frijol mungo, melón, naranja, papaya, frijol de palo (o guandú), granada, arroz y sandía..



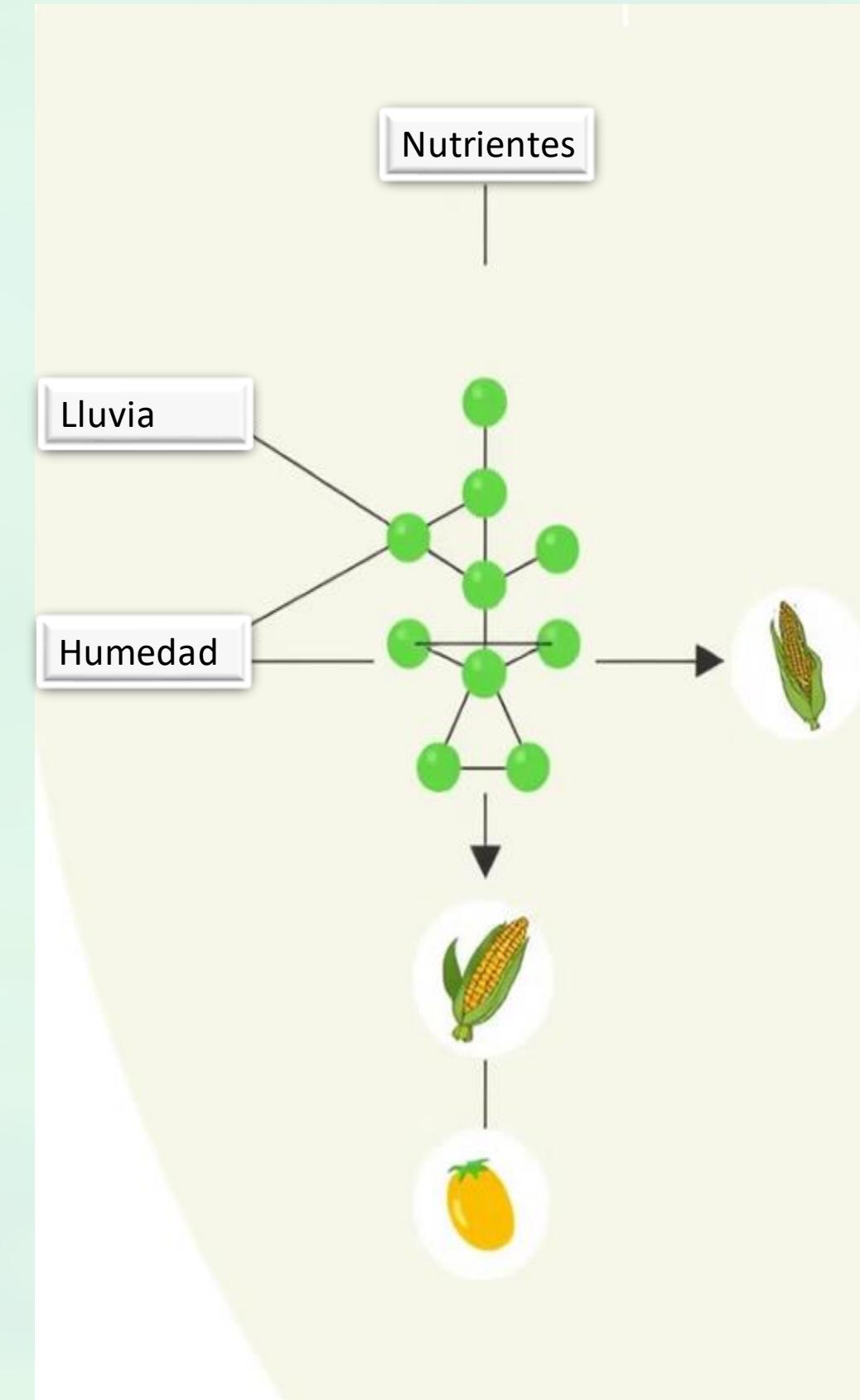
## Tipo de Modelo

Se desarrolló un modelo de clasificación supervisada, donde el algoritmo aprende de un conjunto de datos etiquetados para predecir el cultivo óptimo.



## Parámetros de Entrada

El modelo toma como entradas los parámetros del terreno y del clima, como los niveles de nutrientes del suelo, temperatura, humedad, pH y cantidad de lluvia.



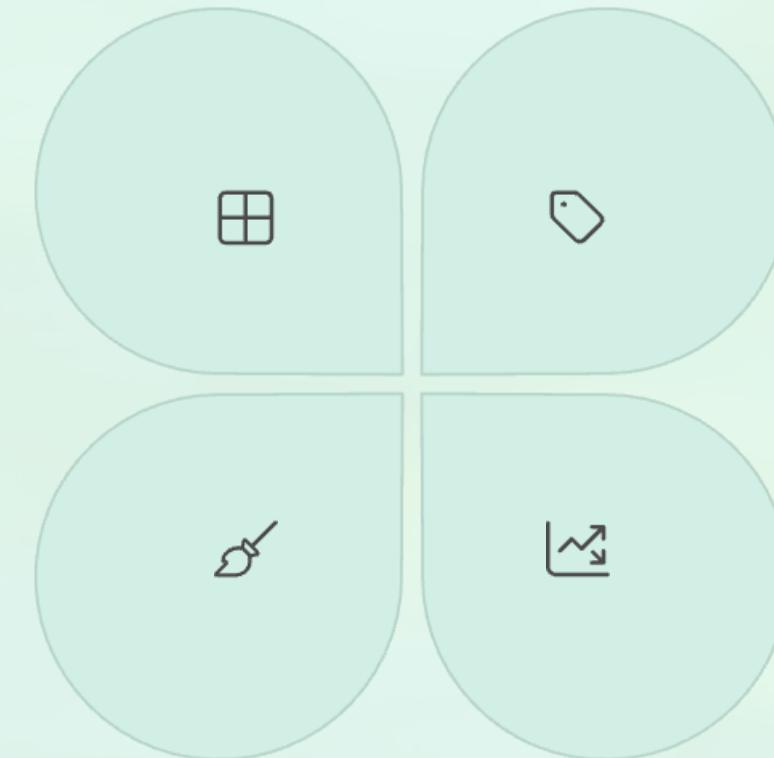
# Preprocesamiento de Datos: Preparando la Información

## División de Datos

Los datos se dividieron estratégicamente en conjuntos de **features** (X), que son las variables de entrada como nutrientes del suelo y clima, y la **etiqueta** (y), que representa el cultivo ideal.

## Ausencia de Nulos

Una ventaja significativa de nuestro dataset es que no se requirió limpieza intensiva, ya que no se encontraron valores nulos ni inconsistencias que pudieran afectar la calidad de los datos.



## Codificación de Etiquetas

Las etiquetas categóricas de los cultivos se transformaron a un formato numérico utilizando **LabelEncoder**, lo cual es esencial para que los algoritmos de Machine Learning puedan procesarlas eficazmente.

## División Entrenamiento/Prueba

Se aplicó una división **train\_test\_split** del 80/20. El 70% de los datos se usó para entrenar el modelo, y el 30% restante para evaluar su rendimiento de manera imparcial.

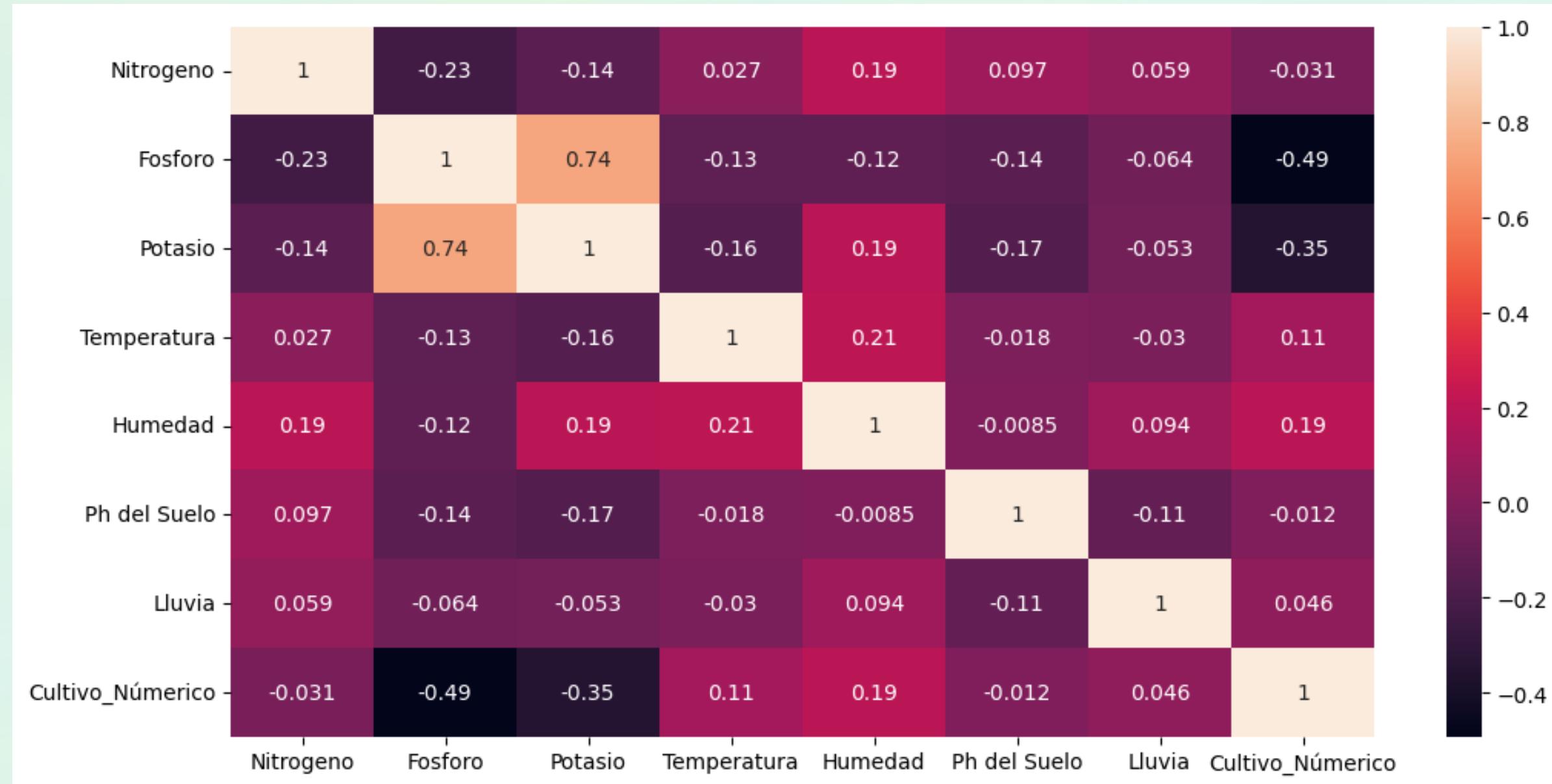
# Preprocesamiento de Datos: Preparando la Información

→

	N	P	K	temperature	humidity	ph	rainfall
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117

#	Column	Non-Null Count	Dtype
---	---	-----	-----
0	N	2200 non-null	int64
1	P	2200 non-null	int64
2	K	2200 non-null	int64
3	temperature	2200 non-null	float64
4	humidity	2200 non-null	float64
5	ph	2200 non-null	float64
6	rainfall	2200 non-null	float64
7	label	2200 non-null	object

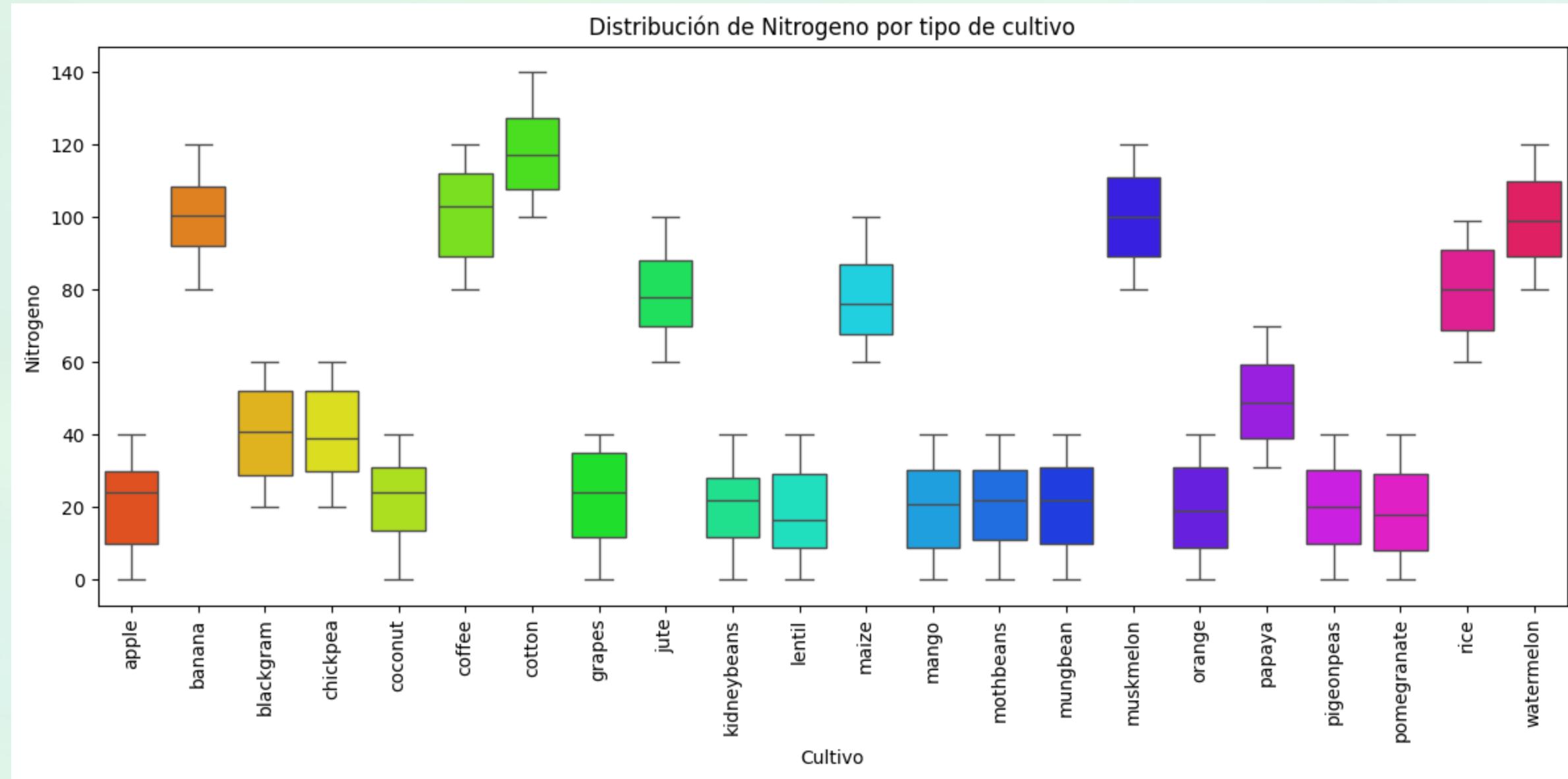
# Preprocesamiento de Datos: Preparando la Información



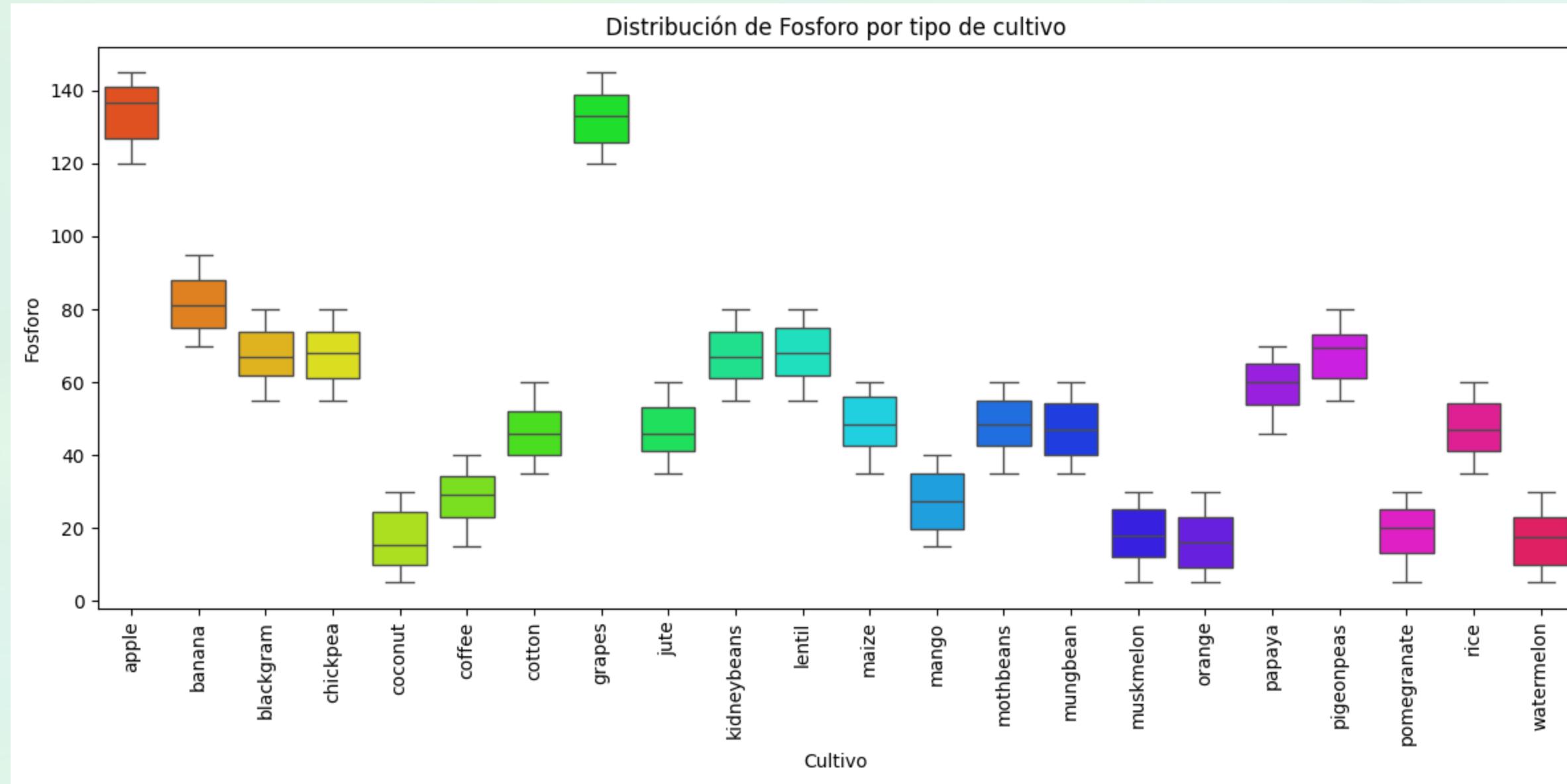
# Preprocesamiento de Datos: Preparando la Información



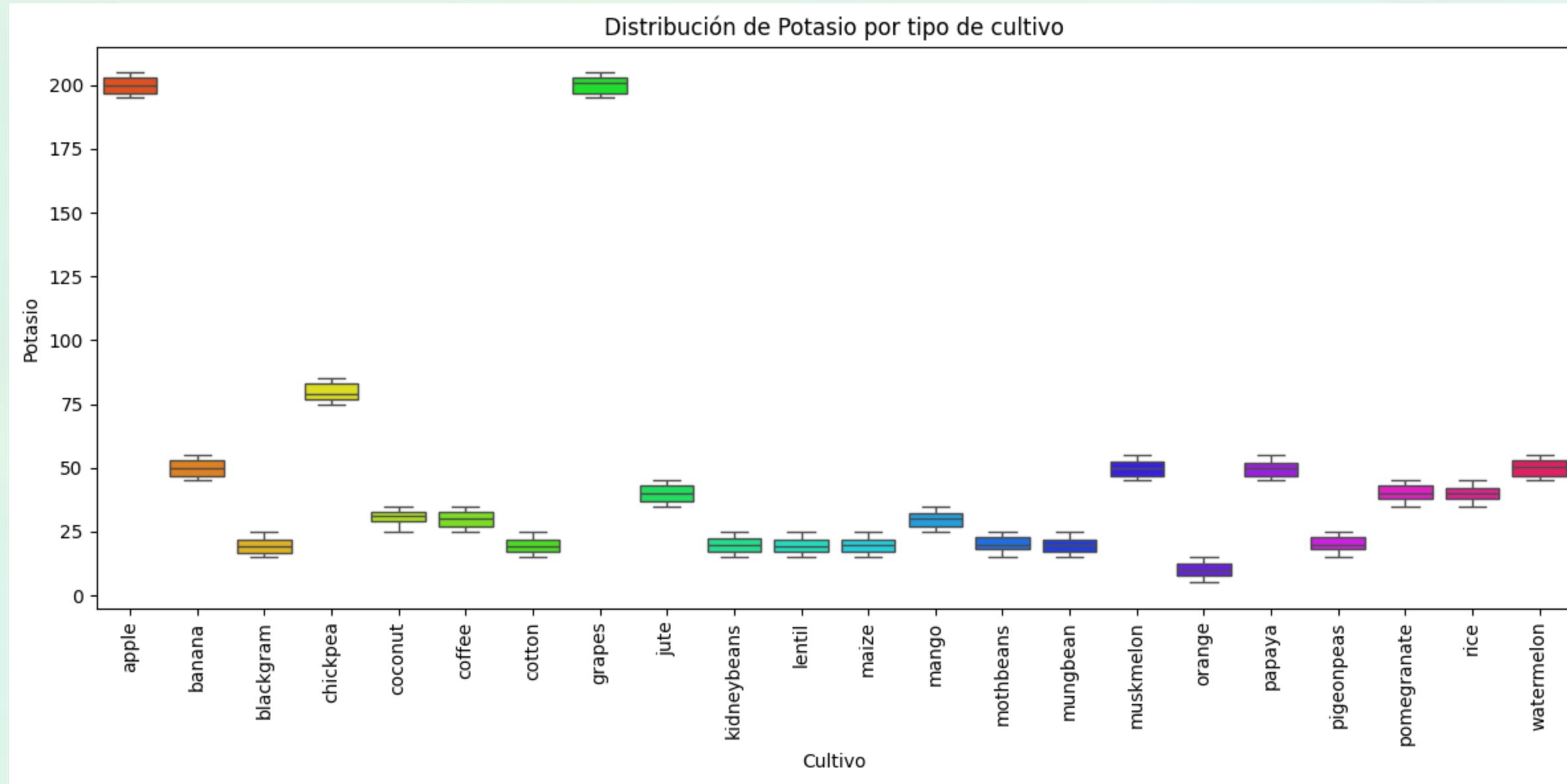
# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



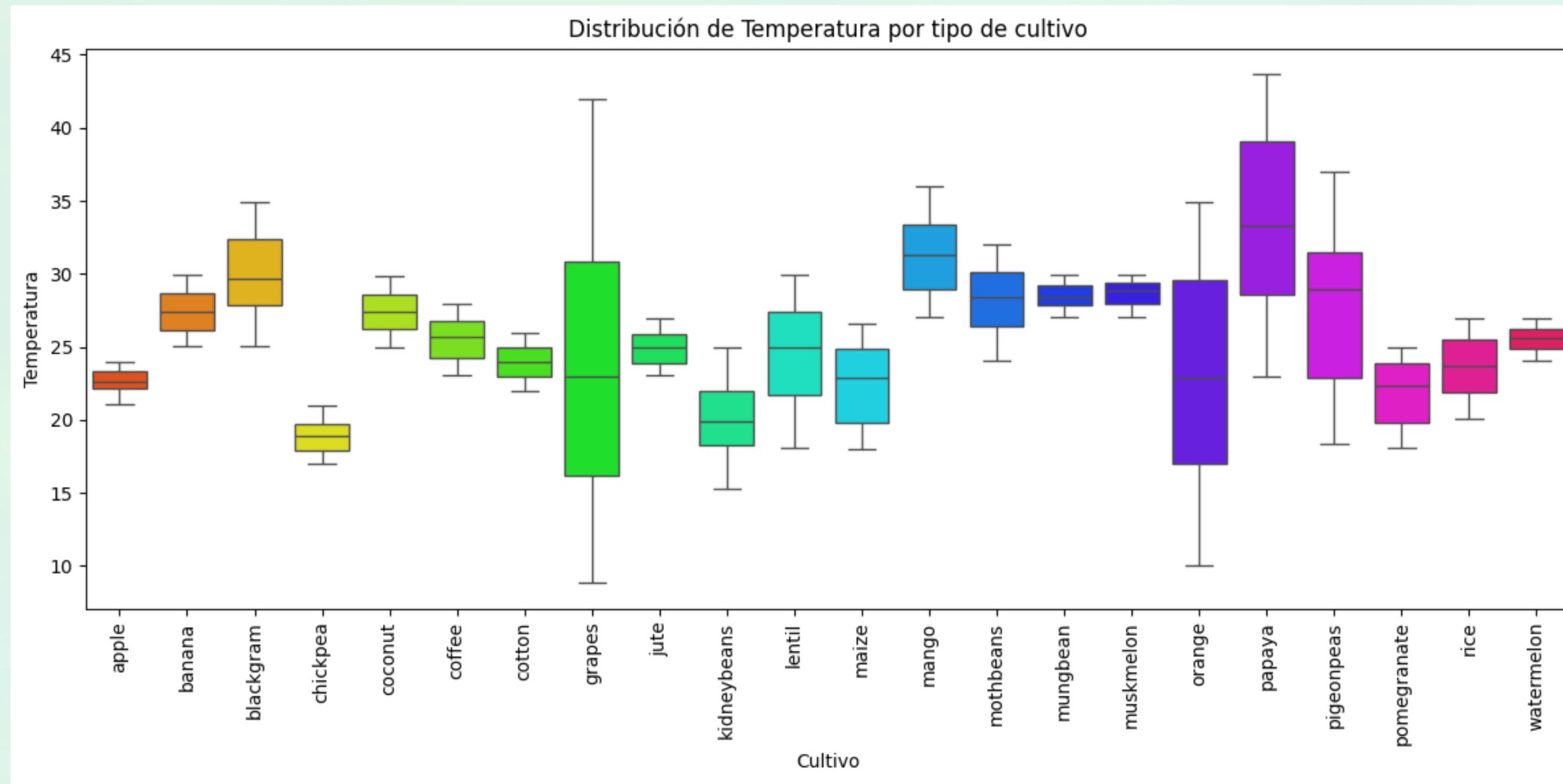
# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



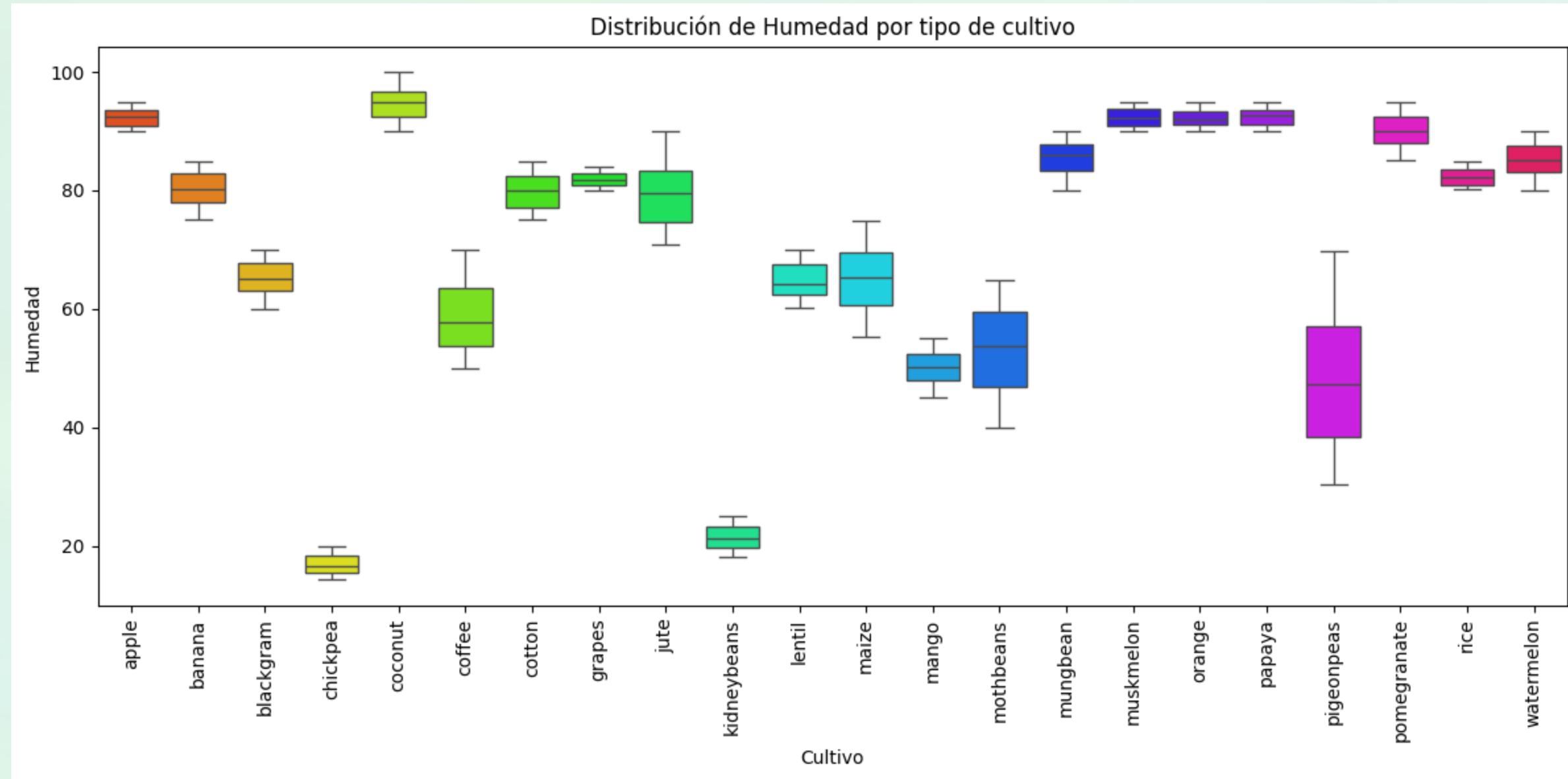
# Preprocesamiento de Datos: Preparando la Información



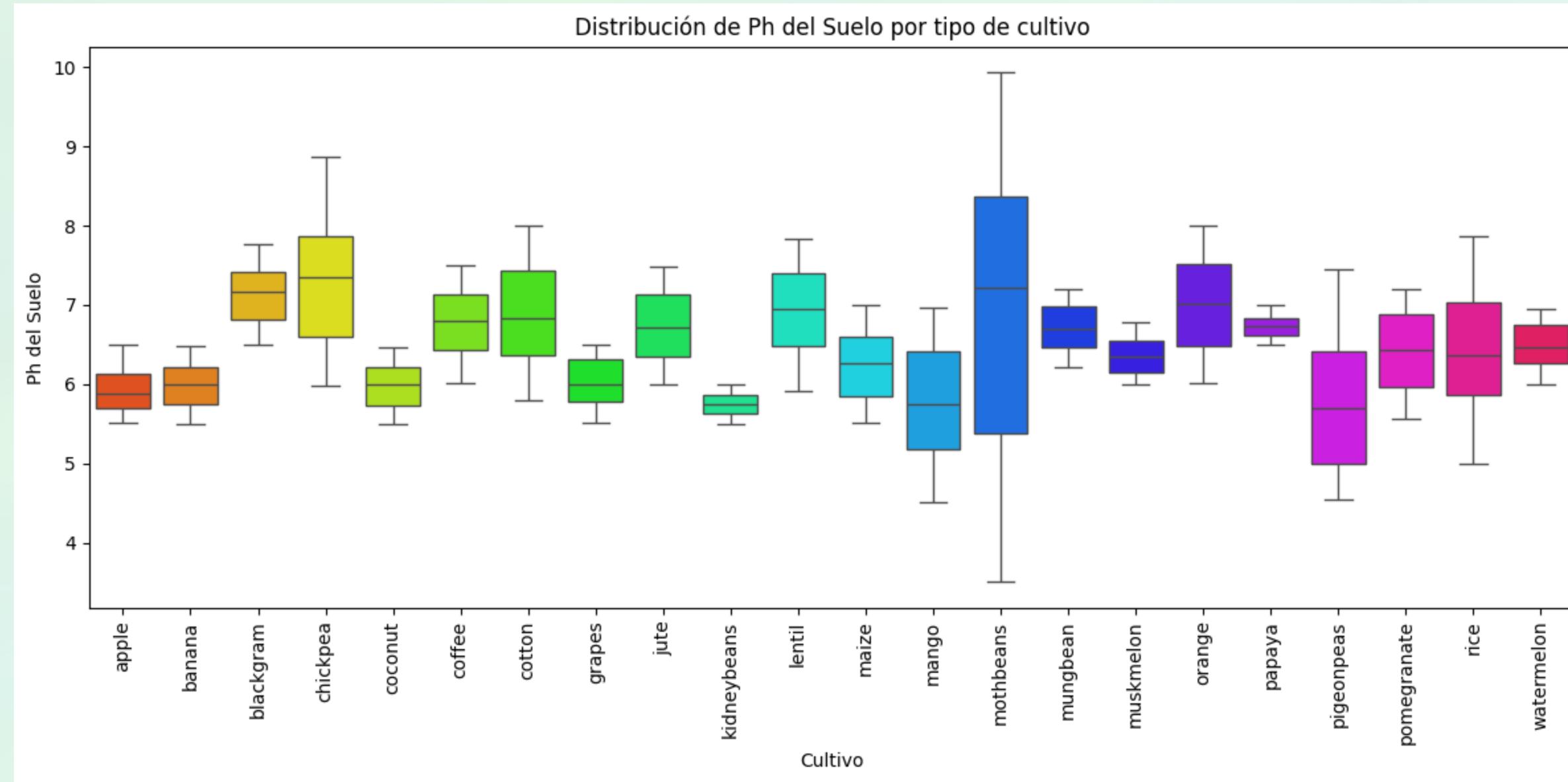
# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



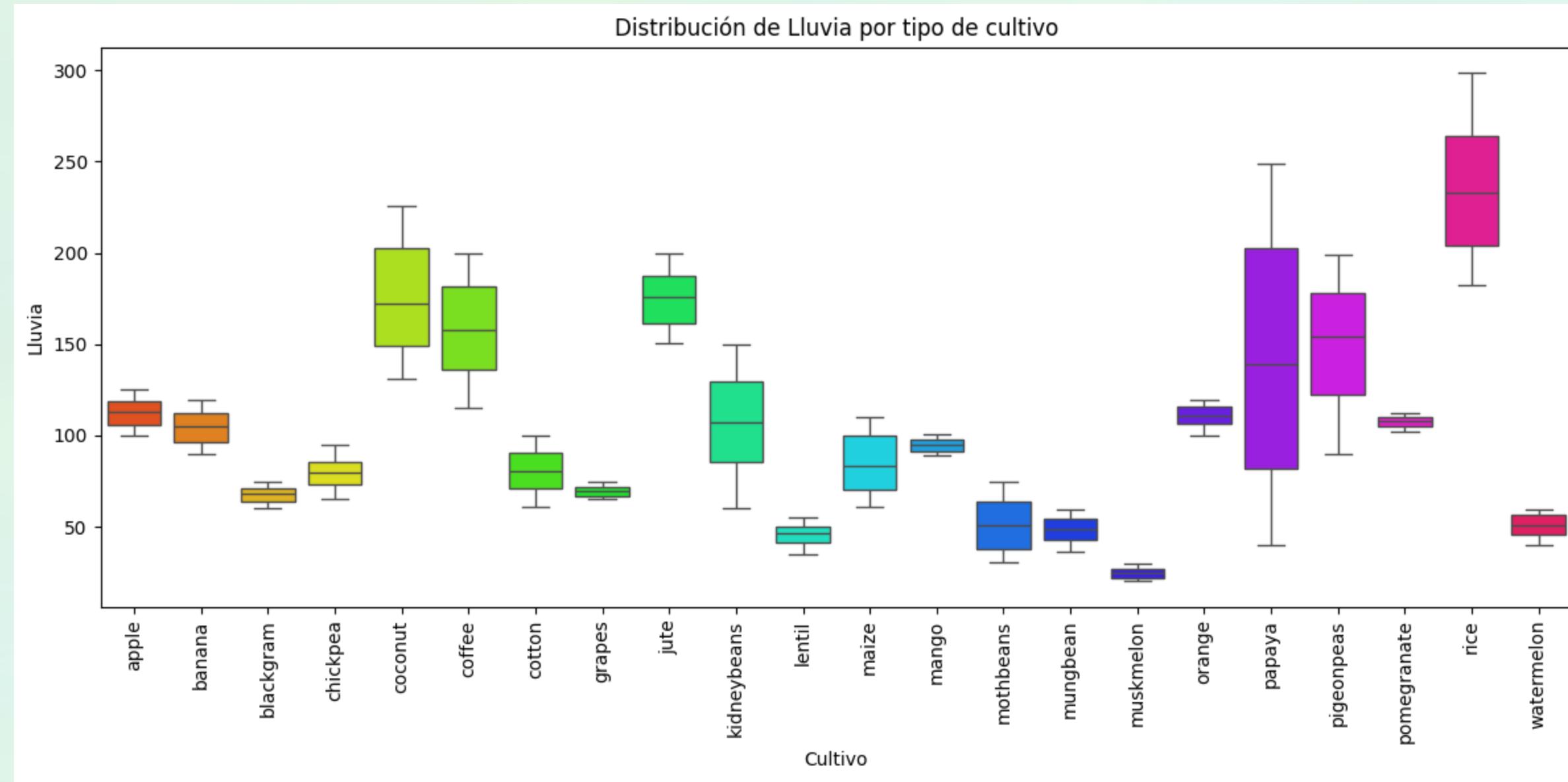
# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



# Preprocesamiento de Datos: Preparando la Información (Análisis Univariado)



# Codificación de Variable categorical (Cultivo)

	Nitrogeno	Fosforo	Potasio	Temperatura	Humedad	Ph del Suelo	Lluvia	Cultivo	Cultivo_Número
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice	20
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice	20
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice	20
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice	20
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice	20

# Codificación de Variable categorical (Cultivo)

0	apple	manzana		
1	banana	banano / plátano	12	mango
2	blackgram	frijol negro pequeño	13	mothbeans
3	chickpea	garbanzo	14	mungbean
4	coconut	coco	15	muskmelon
5	coffee	café	16	orange
6	cotton	algodón	17	papaya
7	grapes	uvas	18	pigeonpeas
8	jute	yute (fibra vegetal)	19	pomegranate
9	kidneybeans	frijoles rojos	20	rice
10	lentil	lenteja	21	watermelon
11	maize	maíz		sandía

# División de Datos:

```
#Selección de la característica o función objetivo y el resto de los datos
X = data.drop(['Cultivo','Cultivo_Númeroico'],axis=1)
y = data['Cultivo_Númeroico']
X_train, X_test, y_train, y_test = train_test_split(X,y,train_size=0.7)
print(X_train.shape,X_test.shape,y_train.shape,y_test.shape)
```

```
(1540, 7) (660, 7) (1540,) (660,)
```



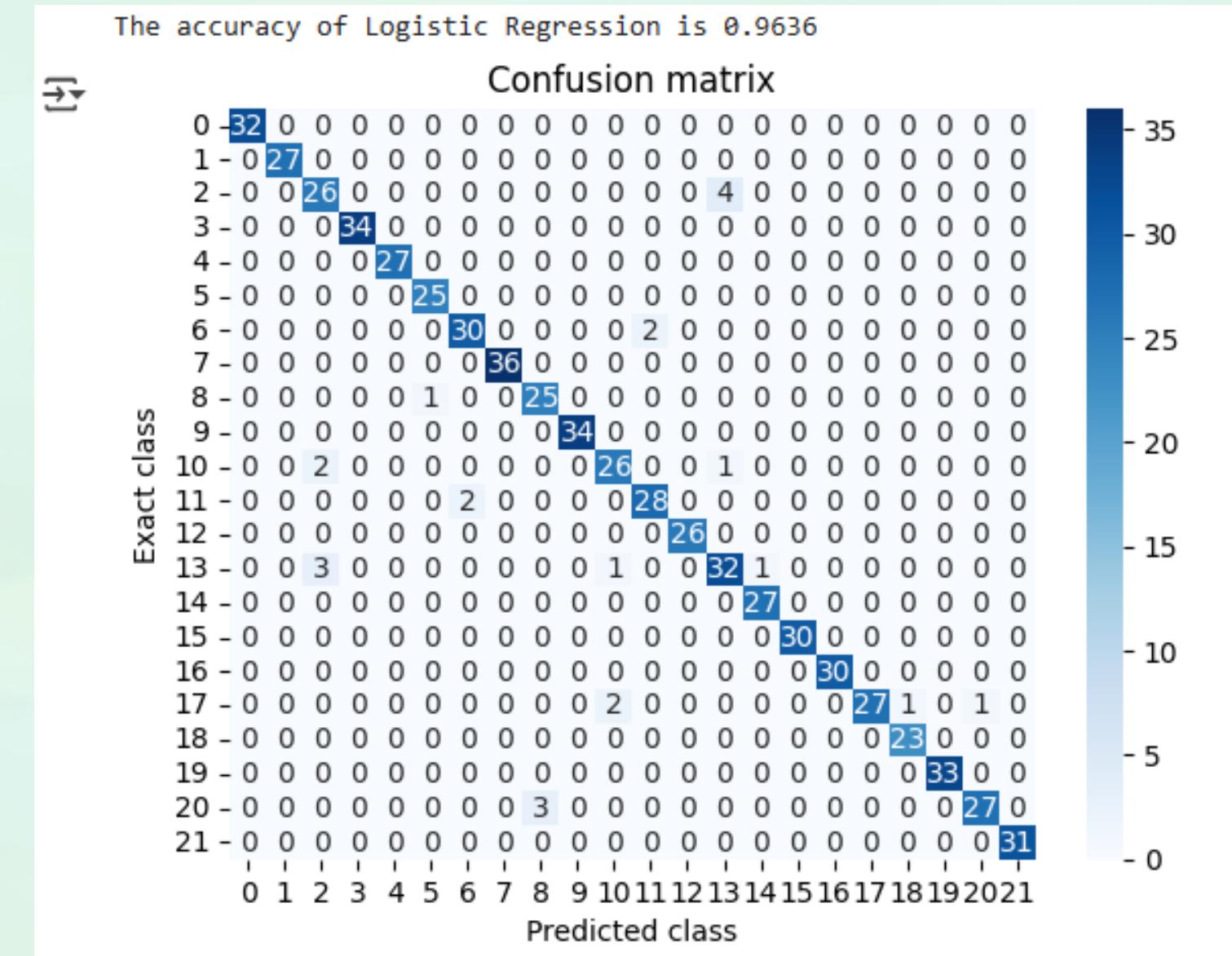
# Algoritmos Evaluados:

accuracy
LogisticRegression 0.963636
KNeighborsClassifier 0.989394
DecisionTreeClassifier 0.981818
RandomForestClassifier 0.995455
GaussianNB 0.996970



# Algoritmos Evaluados:

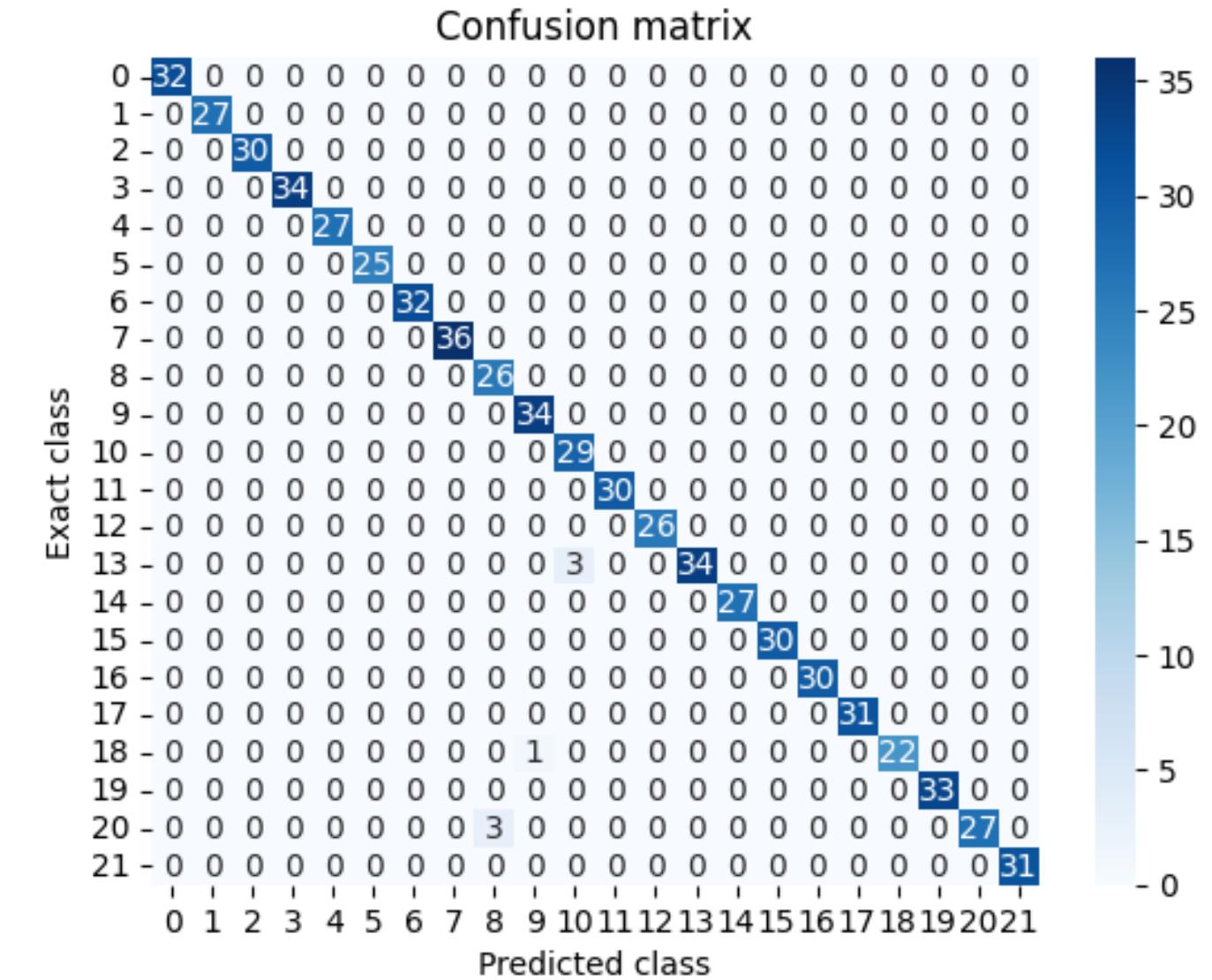
## LogisticRegression





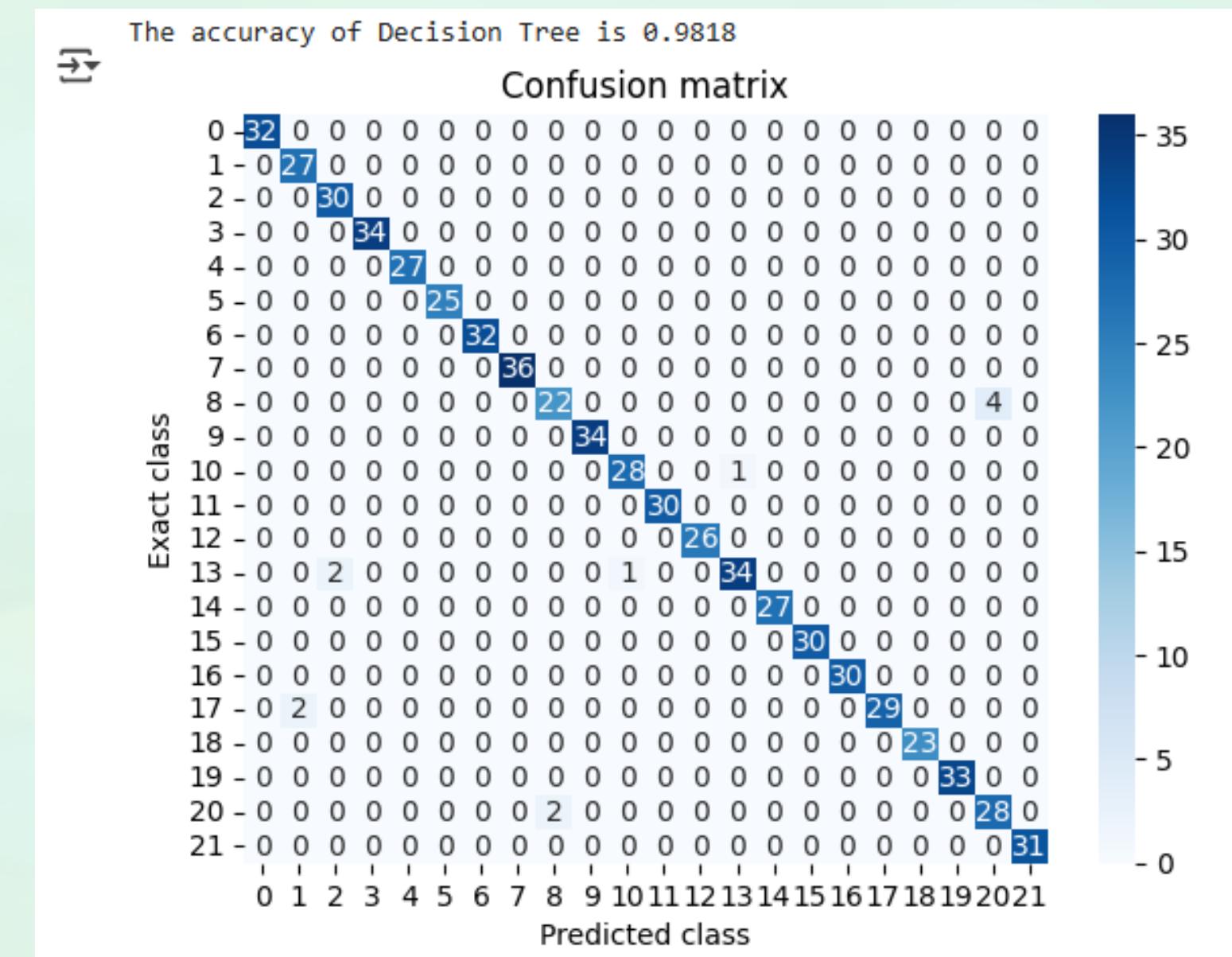
# Algoritmos Evaluados: KNeighborsClassifier

➡ The accuracy of KNeighborsClassifier is 0.9894



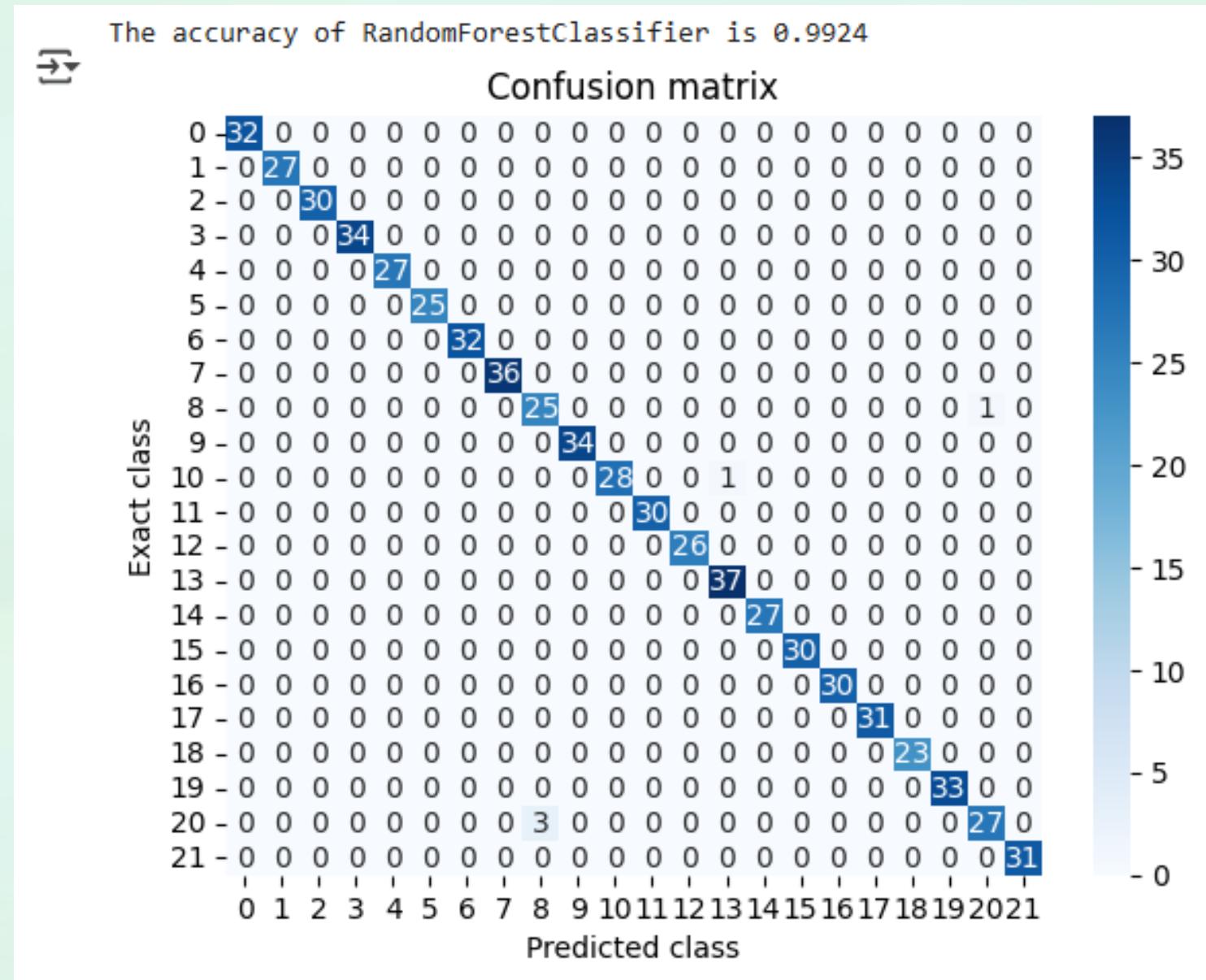


# Algoritmos Evaluados: DecisionTreeClassifier



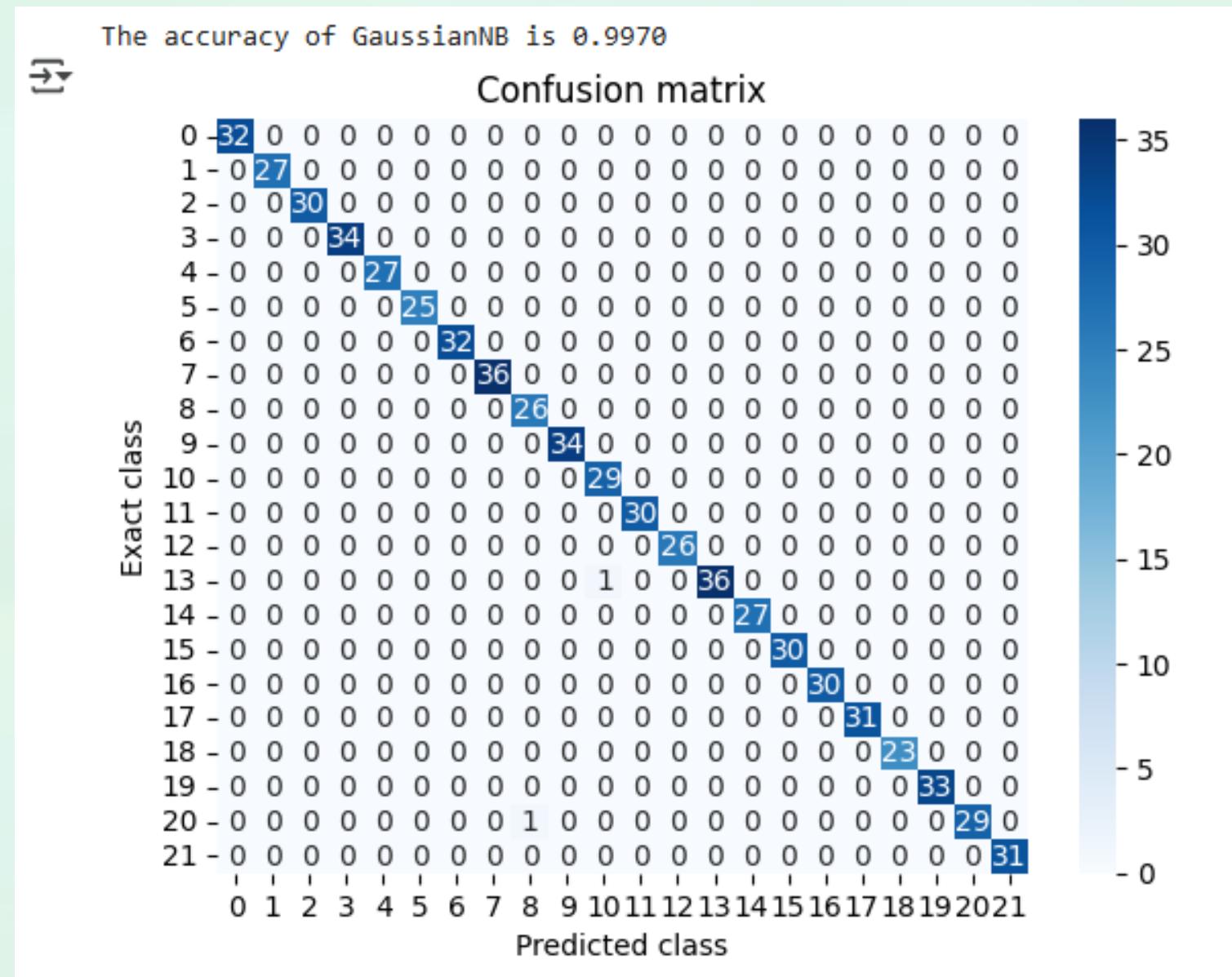


# Algoritmos Evaluados: RandomForestClassifier





# Algoritmos Evaluados: GaussianNB





# Algoritmo Seleccionado: Random Forest Classifier

1

## Clasificación Multiclas

Este algoritmo es excepcionalmente adecuado para problemas de clasificación con múltiples categorías de salida, como la identificación de más de 20 tipos de cultivos.

2

## Robustez al Sobreajuste

Gracias a la combinación de múltiples árboles de decisión. Random Forest minimiza el riesgo de sobreajuste, asegurando un rendimiento consistente en datos no vistos.

3

## Manejo de No Linealidad

Random Forest es capaz de capturar relaciones complejas y no lineales entre las variables de entrada y la variable objetivo, lo cual es crucial en datos agrícolas donde las interacciones son a menudo intrincadas.

# Entrenamiento y Evaluación del Modelo



## Entrenamiento

El modelo **Random Forest Classifier** fue entrenado con un total de 100 árboles de decisión. Esta configuración permite una gran capacidad de aprendizaje y generalización.

## Predictión

Una vez entrenado, el modelo se utilizó para realizar predicciones sobre el conjunto de prueba (el 30% de los datos que no se usaron en el entrenamiento), garantizando una evaluación imparcial de su rendimiento.

## Métricas Clave

Para evaluar la efectividad del modelo, se midieron métricas cruciales como la **precisión (accuracy)**, se generó una **matriz de confusión** detallada, que visualiza los aciertos y errores para cada clase.

# Resultados del Modelo: Predicción Precisa y Confiable

99,5%

## Precisión Total

El modelo alcanzó una impresionante precisión general, demostrando su capacidad para identificar correctamente el cultivo óptimo en una amplia gama de condiciones.

20+

## Cultivos Diferenciados

El algoritmo logró diferenciar con éxito entre todas las variedades de cultivos presentes en el dataset, lo que subraya su robustez y versatilidad para el sector agrícola.

100%

## Exactitud Perfecta

Cultivos específicos como el arroz, el maíz y el plátano fueron predichos con una exactitud del 100%, lo que resalta la capacidad del modelo para dominar ciertas clasificaciones.

Estos resultados confirman que el modelo es una herramienta valiosa para la toma de decisiones agrícolas, ofreciendo un alto nivel de confiabilidad y un gran potencial para mejorar la productividad de manera significativa.



# Conclusiones y Próximos Pasos

## Aplicabilidad Práctica

La alta precisión alcanzada valida la utilidad de esta solución para ser implementada en sistemas de recomendación agrícola en escenarios reales.

¡Gracias por su atención! ¿Preguntas?



# Links

**Kaggle (Dataset):**

<https://www.kaggle.com/datasets/madhuraatmarambhagat/crop-recommendation-dataset>

**Github:**

<https://github.com/JuanPabloArOta/Clasificacion-de-Cultivos>