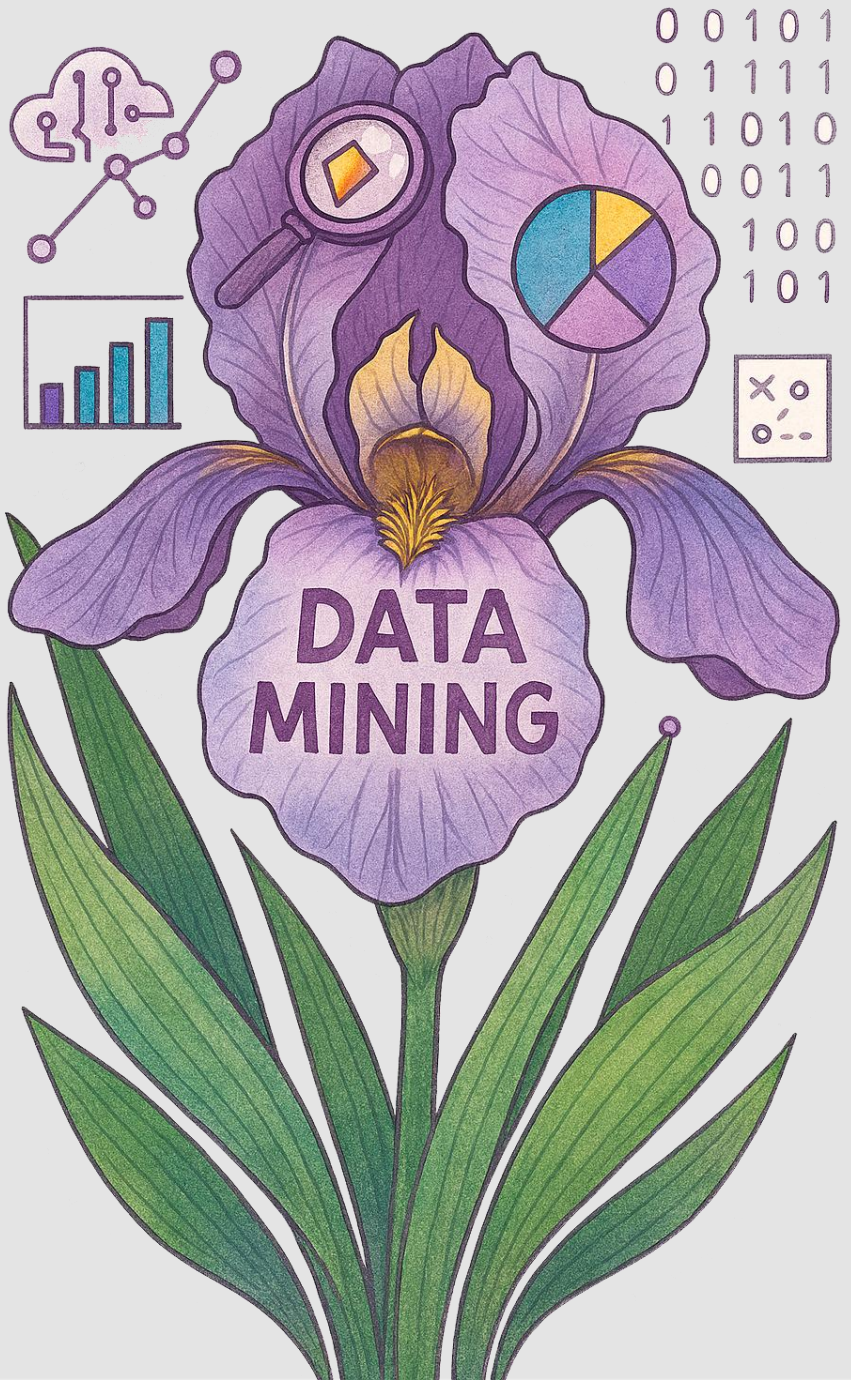
PROYECTO FINAL

IRIS DATASET

JUAN PABLO BORRERO
CARLOS LEAL MEDINA

CORHUILA
CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



¿Cómo se pueden predecir las clases de las flores Iris (Iris-setosa, Iris-versicolor, Iris-virginica) a partir de las características medidas de los sépalos y pétalos?

CORHUILA
CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"

=== ESTADÍSTICOS DESCRIPTIVOS ===

	count	mean	std	min	25%	50%	75%	max
sepal_length	99.0	5.788889	0.831876	4.3	5.1	5.70	6.4	7.9
sepal_width	99.0	3.047475	0.444087	2.0	2.8	3.00	3.3	4.4
petal_length	108.0	3.773148	1.699071	1.0	1.6	4.35	5.1	6.9
petal_width	121.0	1.247934	0.761916	0.1	0.4	1.30	1.8	2.5

```
df.shape
```

```
(150, 5)
```

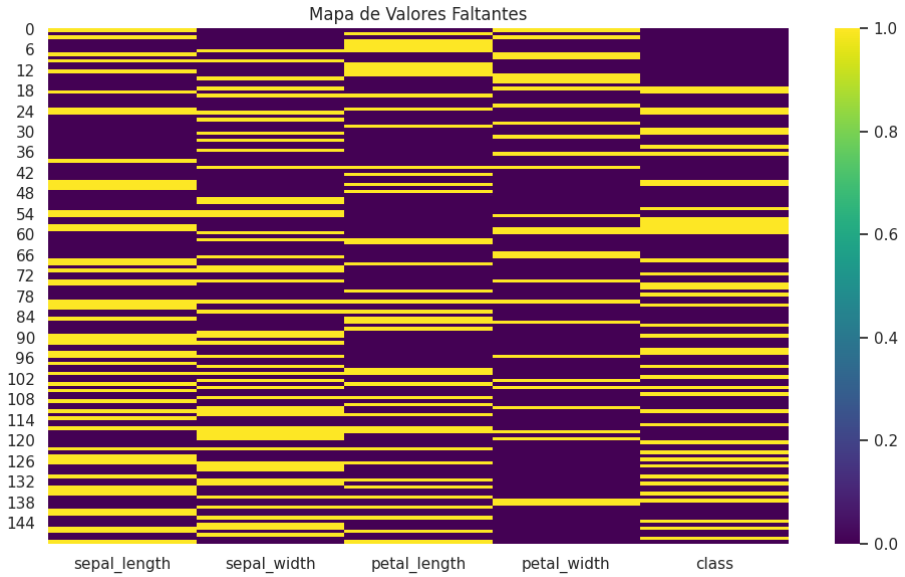
número de registros: 150

Número de variables (columnas): 5

```
=== INFORMACIÓN DEL DATASET ===
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal_length  99 non-null     float64
 1   sepal_width   99 non-null     float64
 2   petal_length  108 non-null    float64
 3   petal_width   121 non-null    float64
 4   class         107 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

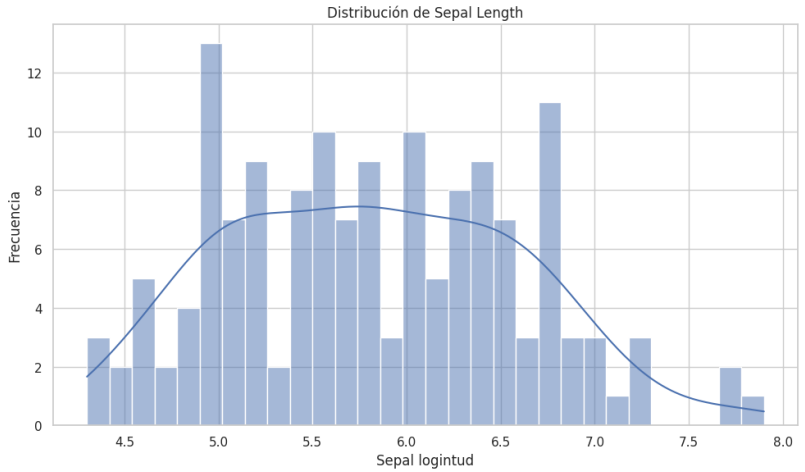
tipos de datos

```
=== VALORES FALTANTES ===
sepal_length    51
sepal_width     51
petal_length    42
petal_width     29
class           43
dtype: int64
```



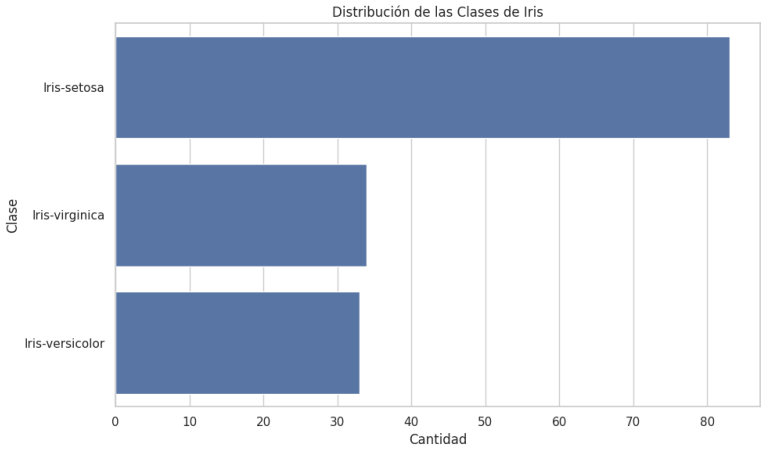
Se identificaron valores faltantes en todas las variables del dataset, siendo sepal_length y sepal_width las más afectadas. Por ello, se realizó un tratamiento de los datos para completar los registros y evitar sesgos en el análisis, lo cual fue confirmado mediante la tabla y la gráfica.

	0
sepal_length	0
sepal_width	0
petal_length	0
petal_width	0
class	0



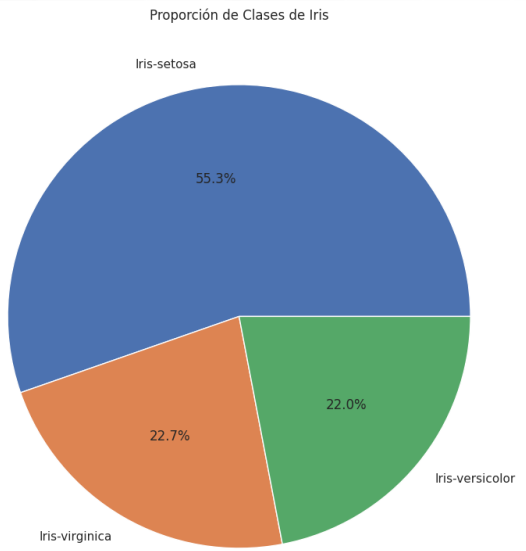
- La distribución es unimodal: la mayor concentración de valores se encuentra aproximadamente entre 5.0 y 6.0 cm.

- la mayor concentración de valores se encuentra aproximadamente entre 5.0 y 5.5 cm, lo que indica una tendencia central clara



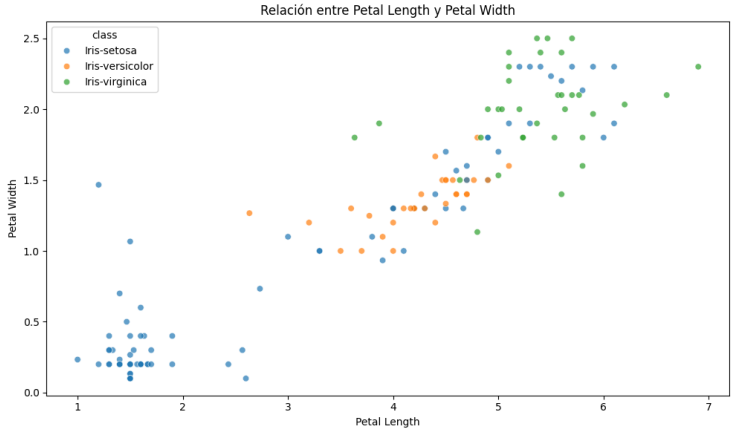
- La clase más representada es Iris-setosa, con una cantidad visiblemente superior respecto a las otras dos clases

- Las clases Iris-virginica y Iris-versicolor tienen cantidades similares



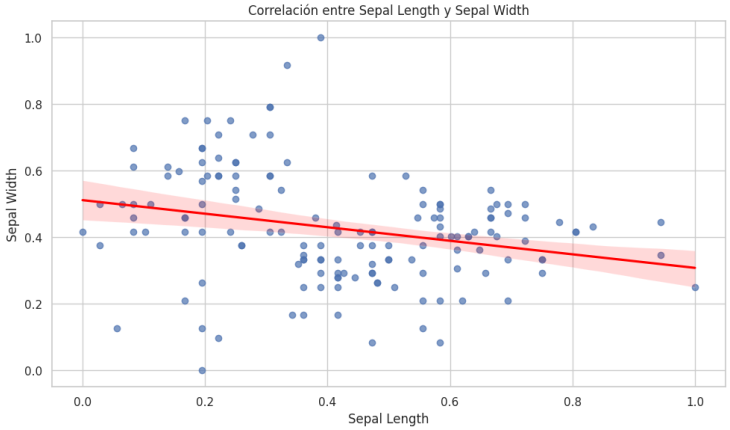
- La clase Iris-setosa representa el 55.3% del total, lo que indica una dominancia significativa en el conjunto de datos

- La visualización en gráfico circular permite identificar rápidamente la proporción relativa



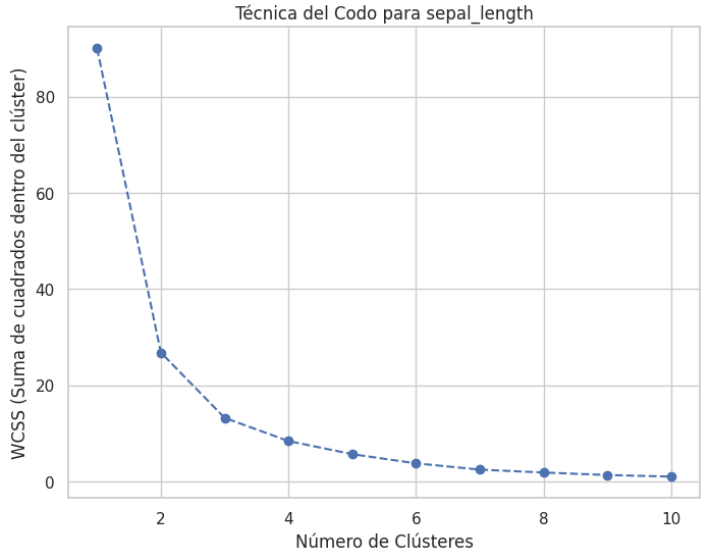
- Muestra la relación entre la longitud del pétalo y el ancho del pétalo en las flores Iris.

- Se observa que a mayor longitud del pétalo, mayor es su ancho, evidenciando una relación positiva.



- Muestran la relación entre la longitud del sépalo y el ancho del sépalo.

- Se observa que, a medida que aumenta la longitud del sépalo, el ancho tiende a disminuir ligeramente.

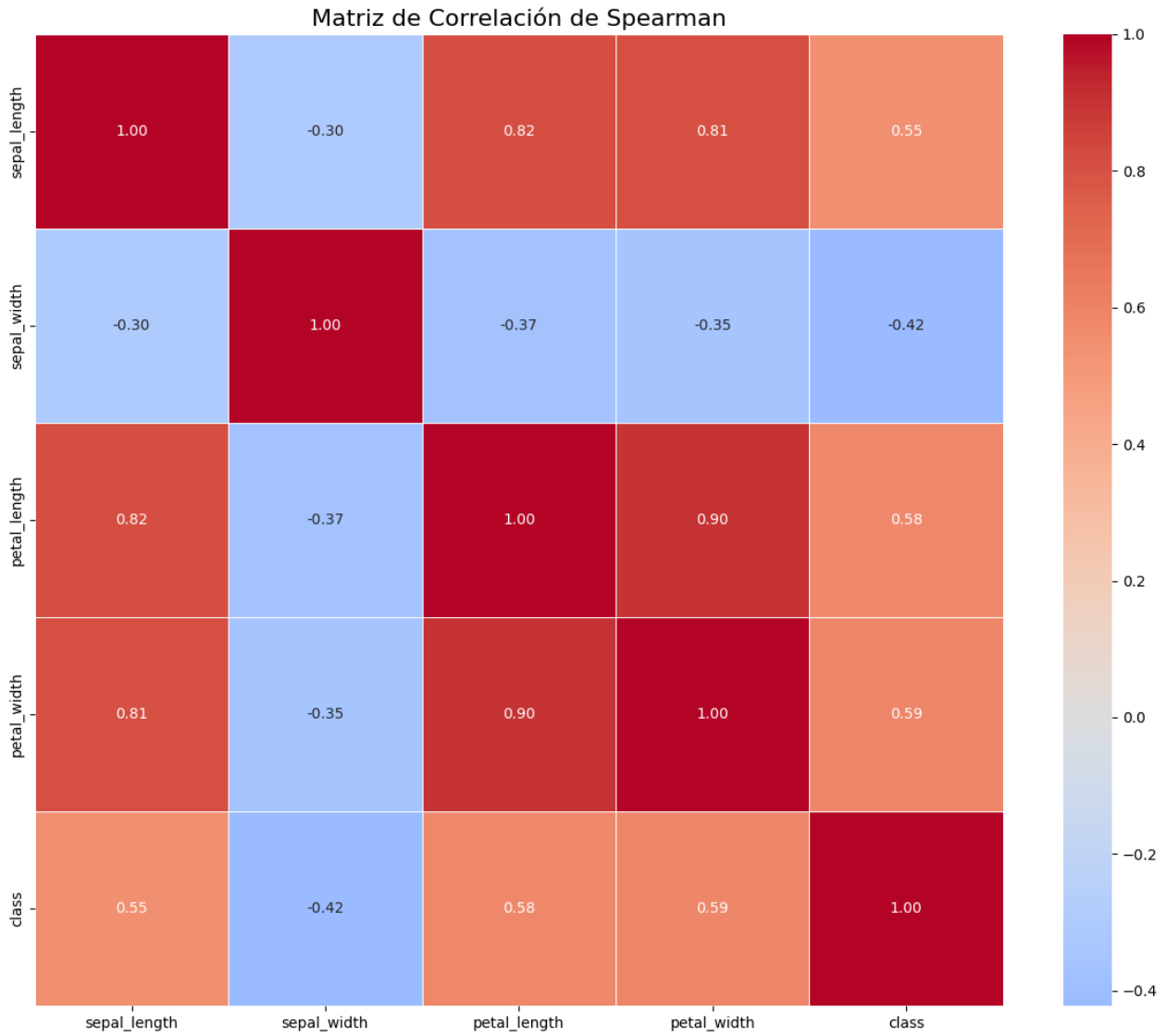


- Se observa una disminución fuerte del WCSS hasta k = 3, donde la curva empieza a estabilizarse.

- Esto indica que 3 clústeres es un número adecuado para agrupar los datos sin perder información relevante.

CORHUILA
CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"



- La correlación más fuerte se presenta entre petal_length y petal_width, con un valor de 0.90, lo que indica una relación muy alta y directa.

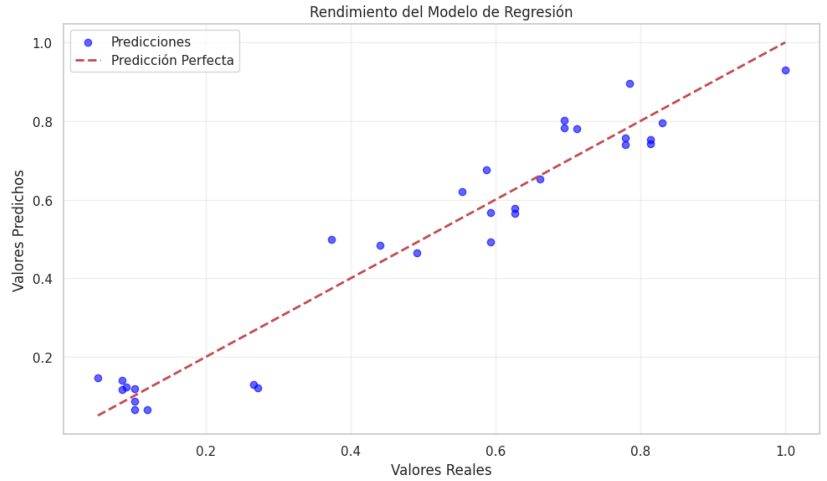
- Esto significa que conociendo la longitud del pétalo es posible estimar con bastante precisión su ancho, ya que ambas variables aumentan casi de forma proporcional.

```
=== EVALUACIÓN DEL MODELO ===
Error Cuadrático Medio (MSE): 0.0070
Raíz del MSE (RMSE): 0.0837
Coeficiente R²: 0.9190

Interpretación R²: El modelo explica el 91.90% de la variabilidad.

=== IMPORTANCIA DE VARIABLES ===
       Variable  Coeficiente
2   petal_width     0.630418
0  sepal_length     0.280413
1   sepal_width    -0.192859
3         class     0.029940
```

70% – 30%
El modelo presenta un error cuadrático medio (MSE) bajo, lo que indica buenas predicciones




```
=== EVALUACIÓN DEL MODELO ===
Error Cuadrático Medio (MSE): 0.0054
Raíz del MSE (RMSE): 0.0732
Coeficiente R²: 0.9347

Interpretación R²: El modelo explica el 93.47% de la variabilidad.

=== IMPORTANCIA DE VARIABLES ===
       Variable  Coeficiente
2   petal_width     0.611505
0  sepal_length     0.299347
1   sepal_width    -0.255007
3         class     0.020790
```

80% - 20%
Un R^2 mayor a 0.93 evidencia un alto poder predictivo

El modelo mostró un **buen desempeño en todas las particiones**, con errores bajos y valores de **R^2 superiores al 91%**, lo que confirma un **alto poder predictivo**.

```
=== EVALUACIÓN DEL MODELO ===
Error Cuadrático Medio (MSE): 0.0066
Raíz del MSE (RMSE): 0.0811
Coeficiente R²: 0.9173

Interpretación R²: El modelo explica el 91.73% de la variabilidad.

=== IMPORTANCIA DE VARIABLES ===
       Variable  Coeficiente
2   petal_width     0.623616
0  sepal_length     0.278878
1   sepal_width    -0.256892
3         class     0.015038
```

90% - 10%
Aunque presenta un desempeño ligeramente inferior a versiones previas, sigue ofreciendo predicciones precisas.

CORHUILA
CORPORACIÓN UNIVERSITARIA DEL HUILA
Vigilada Mineducación

CORPORACIÓN UNIVERSITARIA DEL HUILA - CORHUILA
"Diseño y prestación de servicios de docencia, investigación y extensión de programas de pregrado, aplicando todos los requisitos de las normas ISO implementadas en sus sedes Neiva y Pitalito"