
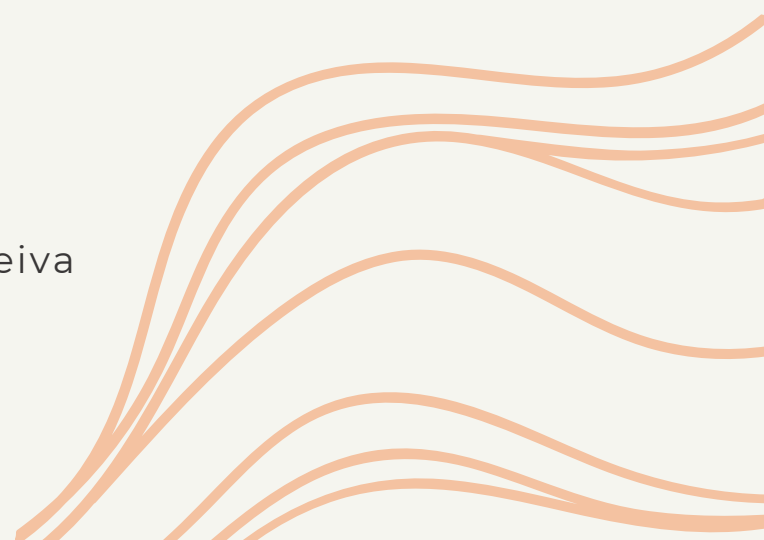



20 de Octubre del  
2024



# LA VISUALIZACIÓN DE DATOS DEL DATASET DE IRIS

Corte 2

Profesor: Sebastián Perdomo Leiva

Escuela Tecnológica Instituto  
Técnico Central

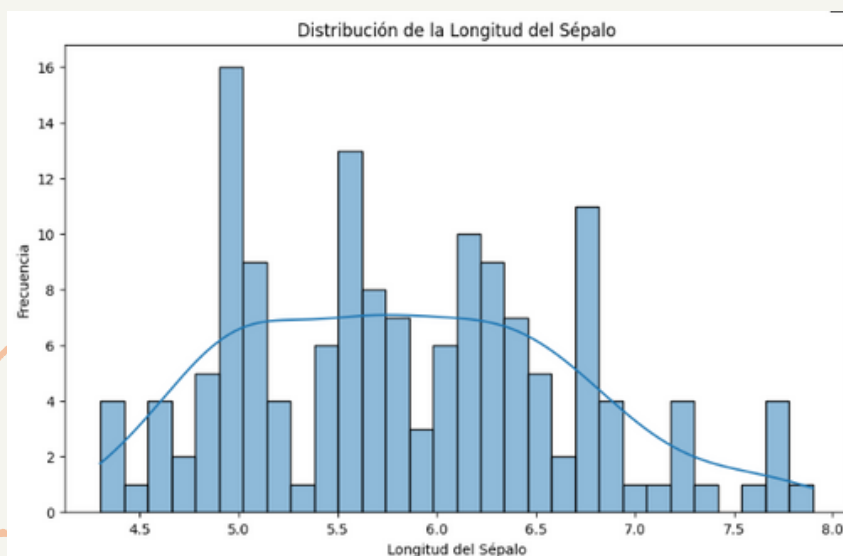
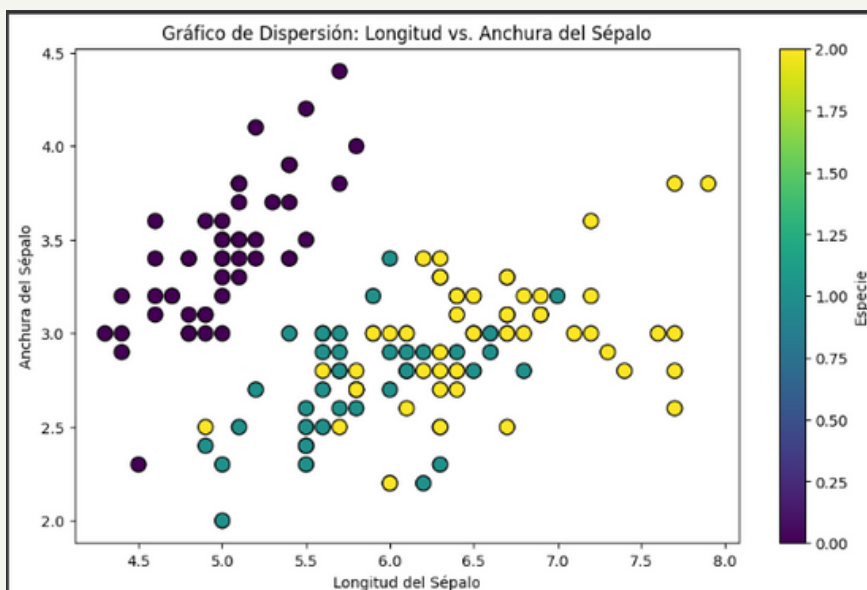
# ÍNDICE

---

🗨	INTRODUCCIÓN .....	1
🗨	ENSAYO .....	2
🗨	CONCLUSIÓN .....	3
🗨	BIBLIOGRAFÍA .....	4

# INTRODUCCIÓN

El Dataset de Iris es uno de los conjuntos de datos más utilizados en el ámbito de la ciencia de datos y el aprendizaje automático. Este dataset contiene mediciones de longitud y anchura del sépalo y el pétalo de tres especies de flores de iris: Setosa, Versicolor y Virginica. A través de diversas visualizaciones, podemos analizar las relaciones entre las características de las flores y cómo estas se agrupan por especie.



# ENSAYO

---

En este análisis, se utilizaron diversas técnicas de visualización para explorar el Dataset de Iris. El primer gráfico, un gráfico de dispersión, representa la longitud y la anchura del sépalo. Este gráfico revela diferencias notables entre las especies de iris. La especie Setosa se agrupa claramente, indicando que tiene características distintivas en comparación con las especies Versicolor y Virginica. Esta visualización es esencial para comprender cómo se separan las especies basadas en sus características físicas.

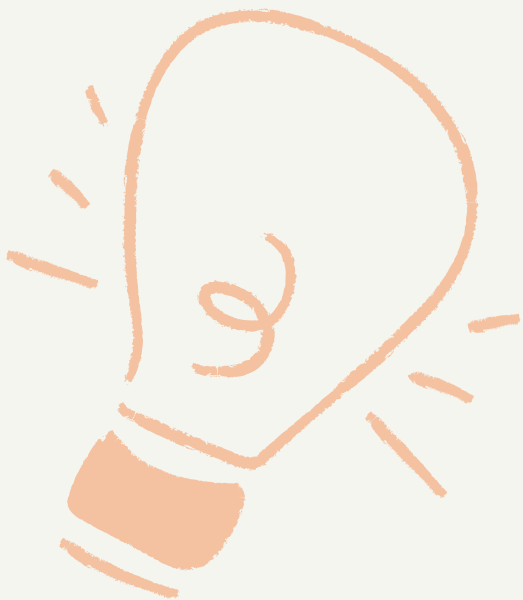
El segundo gráfico es un histograma que muestra la distribución de la longitud del sépalo. La distribución es unimodal, con una concentración significativa de observaciones en el rango de 4 a 7 cm. Esta información es valiosa para entender la variabilidad dentro de las especies y puede servir como base para realizar análisis más profundos. Además, se generó un gráfico de pares, que permite observar todas las combinaciones posibles de las características del dataset. Este gráfico destaca la relación entre la longitud y la anchura del pétalo, mostrando una fuerte correlación entre estas dos variables, especialmente en la especie Setosa. Esta información es crucial para el desarrollo de modelos de clasificación, ya que indica que estas características son útiles para diferenciar entre las especies.

Finalmente, se creó un mapa de calor que muestra la matriz de correlación de las variables del dataset. Este gráfico resalta la alta correlación positiva entre la longitud y la anchura del pétalo, sugiriendo que un aumento en la longitud del pétalo también está asociado con un aumento en su anchura. Este tipo de análisis es fundamental para los científicos de datos, ya que les ayuda a identificar variables que pueden influir en la clasificación y predicción de especies.

# CONCLUSIÓN

---

La visualización de datos en el análisis del Dataset de Iris demuestra su importancia en la comprensión y la interpretación de patrones complejos. A través de gráficos de dispersión, histogramas, gráficos de pares y mapas de calor, se puede obtener una visión clara de las relaciones entre las características de las flores y cómo estas se agrupan por especie. Estas visualizaciones no solo facilitan la identificación de tendencias y correlaciones, sino que también son herramientas esenciales para la toma de decisiones en el ámbito de la biología y la ciencia de datos.



# BIBLIOGRAFÍA

---

- Fisher, R.A. (1936). "The use of multiple measurements in taxonomic problems." *Annals of Eugenics*, 7(2), 179-188.
- Seaborn Documentation. (2021). "Seaborn: statistical data visualization." <https://seaborn.pydata.org/>
- Matplotlib Documentation. (2021). "Matplotlib: plotting with Python." <https://matplotlib.org/>
- Universidad de los Andes. (2020). "Análisis Exploratorio de Datos y Visualización." <https://www.uniandes.edu.co>
- 
-