1 de abril de 2025

**Actividad 1**

Actividad de Aprendizaje 6 – Ciencia de Datos con Python

ANÁLISIS Y DESARROLLO DE SOFTWARE

eSTEFANÍA BETANCUR LEÓN

ANJHELL REYES CUFIÑO

SERGIO MORALES

ANDY RIOS

**1. Introducción**

Utilizando un conjunto de datos sencillo (Vamos a utilizar en el ejemplo el conjunto Iris de la librería Seaborn); como la estructura definida para ‘la rueda de la vida’ de Industrias Coquito Amarillo (Ver guía de actividad 1), se ha realizado un análisis exploratorio de los datos (EDA, por sus siglas en inglés). En este análisis se identifican patrones, valores atípicos y se presentan algunas visualizaciones básicas, como gráficas de barras e histogramas.

**2. Histograma - Largo del Sépalo**

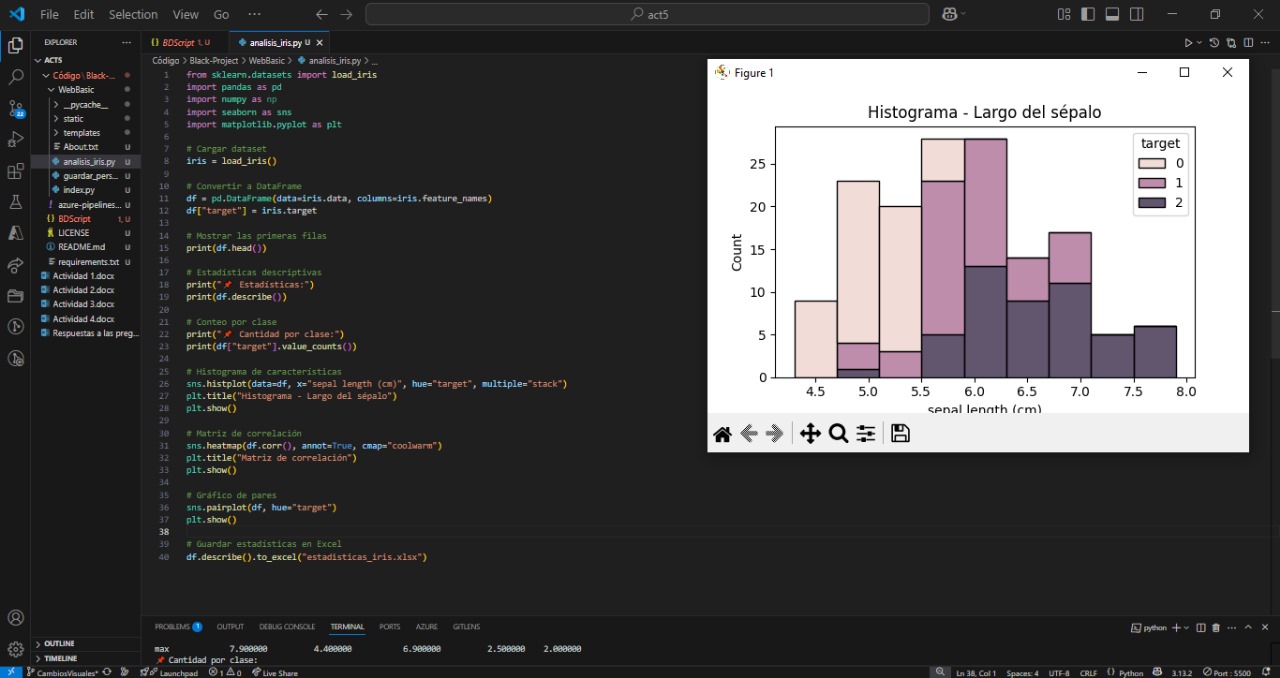
El histograma presentado muestra la distribución del largo del sépalo para las tres clases de flores en el conjunto de datos. Se puede observar lo siguiente:

- Las clases presentan distribuciones diferenciadas, aunque hay solapamiento entre ellas.

- La mayoría de los valores de largo del sépalo oscilan entre 4.5 cm y 7.5 cm.

- La clase 0 (Setosa) tiende a tener sépalos más cortos en comparación con las otras dos clases.

- Las clases 1 (Versicolor) y 2 (Virginica) presentan valores superpuestos, aunque la clase 2 suele tener valores mayores.



**3. Matriz de Correlación**

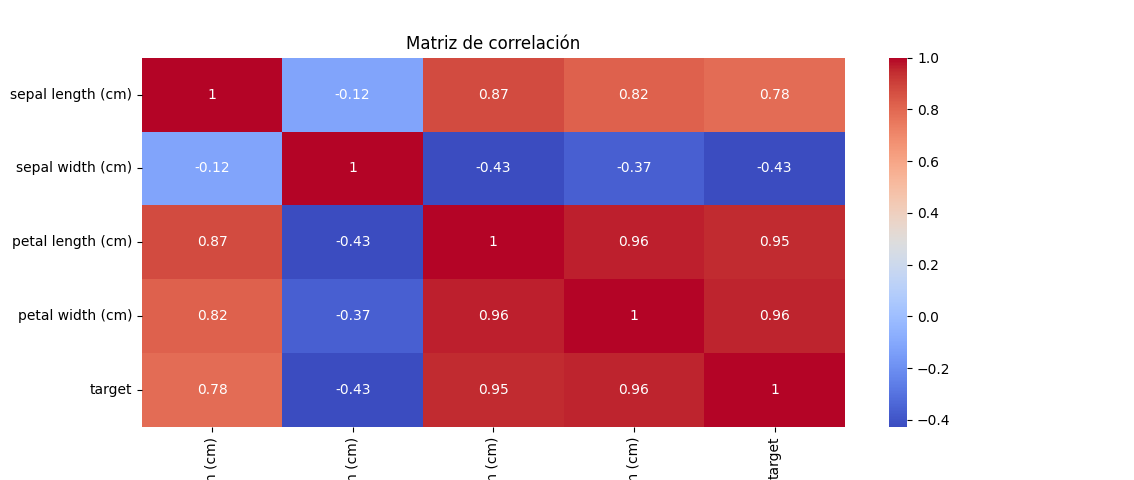
El análisis de la matriz de correlación revela los siguientes puntos clave:

- Existe una fuerte correlación positiva entre el largo del pétalo y el ancho del pétalo (0.96), lo que indica que a medida que una de estas dimensiones aumenta, la otra también tiende a hacerlo.

- La variable **target** muestra una correlación positiva significativa con el largo del pétalo (0.95) y el ancho del pétalo (0.96), lo que sugiere que estas variables son determinantes para la clasificación de las flores.

- La correlación entre el largo del sépalo y el ancho del sépalo es baja (-0.12), lo que indica que estas variables no están directamente relacionadas.

- El ancho del sépalo presenta una correlación negativa con la mayoría de las otras características, lo que podría indicar que su influencia en la clasificación de las especies es menor.



**4. Conclusiones**

El análisis exploratorio de datos nos ha permitido extraer información valiosa sobre la distribución y relaciones de las variables en el conjunto de datos Iris. La visualización del histograma nos muestra una diferenciación clara entre las especies en función del largo del sépalo, mientras que la matriz de correlación resalta la importancia del largo y ancho del pétalo en la clasificación de las flores. Estos hallazgos podrían ser utilizados para construir modelos de clasificación más precisos basados en aprendizaje automático.

