

Actividad de Aprendizaje 6-Análisis de datos(EDA)

Mónica Fernanda Rojas Celis

Yeimy Vanessa Ricardo Ballestas

Daniel Gómez Pérez

Servicio Nacional de Aprendizaje

Tecnología en Análisis y desarrollo de software – Ficha 2828523

Medellín, 10 de abril de 2025

Actividad 1-EDA

1. Instalación de bibliotecas necesarias

pip install pandas numpy matplotlib seaborn

```
danie@LAPTOP-0MT3MCT1 MINGW64 /c/Users/danie/Desktop/CoquitoWeb
$ pip install pandas numpy matplotlib seaborn
Requirement already satisfied: pandas in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (2.2.3)
Requirement already satisfied: numpy in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (2.0.2)
Requirement already satisfied: matplotlib in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (3.9.4)
Requirement already satisfied: seaborn in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (0.13.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (from pandas) (2025.2)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (from pandas) (2.9.0)
Requirement already satisfied: contourpy>=1.0.1 in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (from matplotlib) (1.3.0)
Requirement already satisfied: pillow>=8 in c:\users\danie\appdata\local\programs\python\python39\lib\site-packages (from matplotlib) (11.1.0)
```

2. Importación de librerías y carga de datos

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Cargar el conjunto de datos Iris desde seaborn
#(aquí deberá cargar su conjunto de datos)
data = sns.load_dataset('iris')

# Ver las primeras filas del conjunto de datos
print(data.head())
```

```
app.py > ...
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 # Cargar el conjunto de datos Iris desde seaborn
6 #(aquí deberá cargar su conjunto de datos)
7 data = sns.load_dataset('iris')
8 # Ver las primeras filas del conjunto de datos
9 print(data.head())
```

```

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
• $ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa

```

3. Descripción general del conjunto de datos

Descripción general del conjunto de datos

```
print(data.describe())
```

```

10
11 # Descripción general del conjunto de datos
12 print(data.describe())

```

```

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
• $ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
• $ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa
  sepal_length  sepal_width  petal_length  petal_width

```

```

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
$ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa
count    150.000000    150.000000    150.000000    150.000000
mean       5.843333     3.057333     3.758000     1.199333
std        0.828066     0.435866     1.765298     0.762238
min        4.300000     2.000000     1.000000     0.100000
25%        5.100000     2.800000     1.600000     0.300000
50%        5.800000     3.000000     4.350000     1.300000
75%        6.400000     3.300000     5.100000     1.800000
max        7.900000     4.400000     6.900000     2.500000

```

4. Identificación de valores atípicos

Generar boxplots para cada una de las características del conjunto de datos

```
plt.figure(figsize=(12, 8))
```

Boxplot de todas las columnas numéricas

```
sns.boxplot(data=data[['sepal_length', 'sepal_width',
'petal_length', 'petal_width']])
```

```
plt.title("Boxplot de las características del Iris")
```

```
plt.show()
```

```

14 # Generar boxplots para cada una de las características del conjunto de datos
15 plt.figure(figsize=(12, 8))
16 # Boxplot de todas las columnas numéricas
17 sns.boxplot(data=data[['sepal_length', 'sepal_width',
18 'petal_length', 'petal_width']])
19 plt.title("Boxplot de las características del Iris")
20 plt.show()

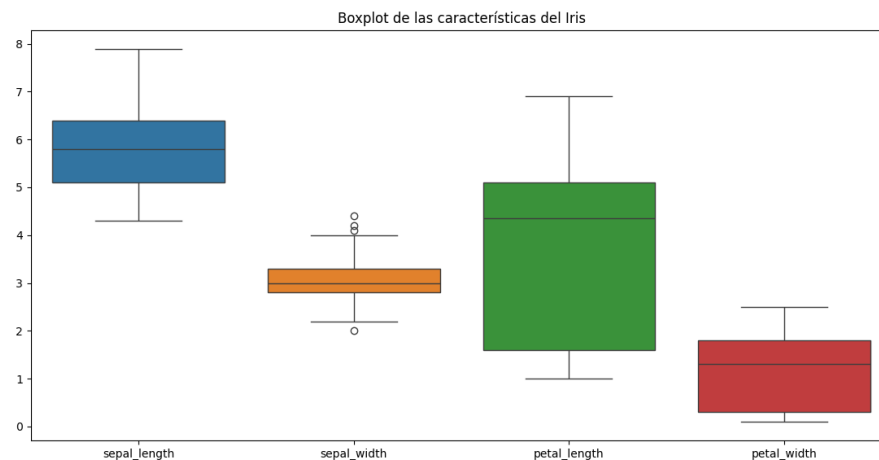
```

```

$ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa

  sepal_length  sepal_width  petal_length  petal_width
count    150.000000    150.000000    150.000000    150.000000
mean       5.843333     3.057333     3.758000     1.199333
std        0.828066     0.435866     1.765298     0.762238
min        4.300000     2.000000     1.000000     0.100000
25%        5.100000     2.800000     1.600000     0.300000
50%        5.800000     3.000000     4.350000     1.300000
75%        6.400000     3.300000     5.100000     1.800000
max        7.900000     4.400000     6.900000     2.500000

```



5. Visualización de la distribución de los datos (Histogramas)

Histograma de las características numéricas

```
data[['sepal_length', 'sepal_width', 'petal_length',
      'petal_width']].hist(bins=15, figsize=(12, 8))
```

```
plt.suptitle("Histogramas de las características del Iris")
```

```
plt.show()
```

```

21
22 # Histograma de las características numéricas
23 data[['sepal_length', 'sepal_width', 'petal_length',
24       'petal_width']].hist(bins=15, figsize=(12, 8))
25 plt.suptitle("Histogramas de las características del Iris")
26 plt.show()

```

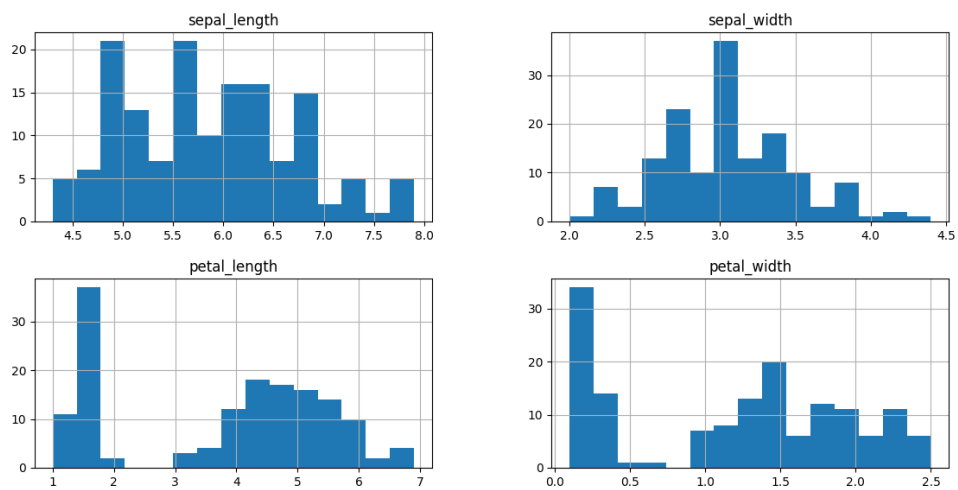
```

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
$ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa

  sepal_length  sepal_width  petal_length  petal_width
count    150.000000    150.000000    150.000000    150.000000
mean       5.843333     3.057333     3.758000     1.199333
std        0.828066     0.435866     1.765298     0.762238
min        4.300000     2.000000     1.000000     0.100000
25%        5.100000     2.800000     1.600000     0.300000
50%        5.800000     3.000000     4.350000     1.300000
75%        6.400000     3.300000     5.100000     1.800000
max        7.900000     4.400000     6.900000     2.500000

```

Histogramas de las características del Iris



6. Relación entre las características (Gráficas de dispersión)

Pairplot para ver la relación entre las características numéricas

```
sns.pairplot(data, hue='species')
```

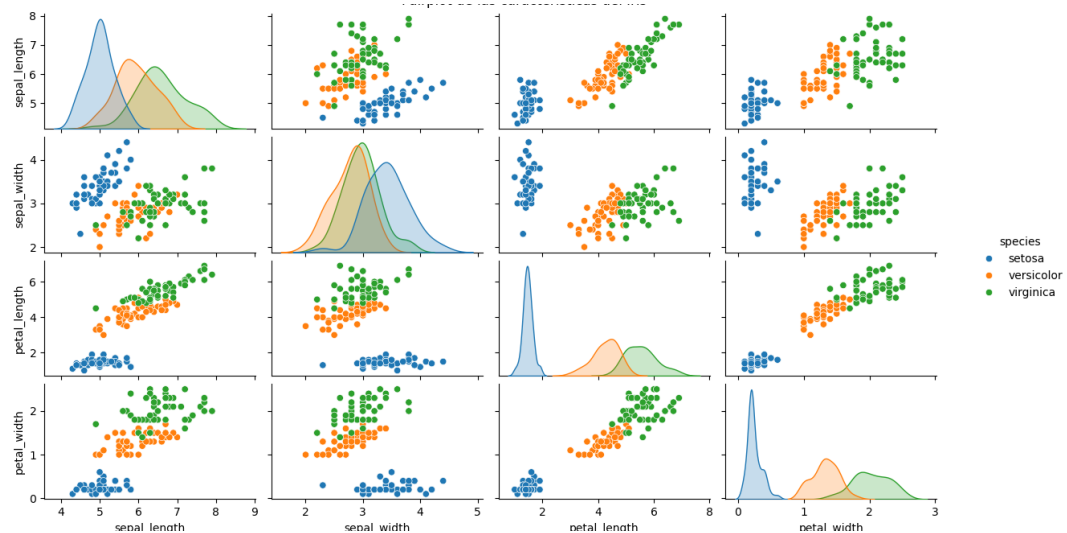
```
plt.suptitle("Pairplot de las características del Iris", y=1.02)
```

```
plt.show()
```

```

danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
$ python app.py
  sepal_length  sepal_width  petal_length  petal_width  species
0           5.1           3.5           1.4           0.2  setosa
1           4.9           3.0           1.4           0.2  setosa
2           4.7           3.2           1.3           0.2  setosa
3           4.6           3.1           1.5           0.2  setosa
4           5.0           3.6           1.4           0.2  setosa
count      150.000000    150.000000    150.000000    150.000000
mean        5.843333     3.057333     3.758000     1.199333
std         0.828066     0.435866     1.765298     0.762238
min         4.300000     2.000000     1.000000     0.100000
25%         5.100000     2.800000     1.600000     0.300000
50%         5.800000     3.000000     4.350000     1.300000
75%         6.400000     3.300000     5.100000     1.800000
max         7.900000     4.400000     6.900000     2.500000

```



7. Correlación entre características

Calcular la matriz de correlación

```
correlation_matrix = data.corr()
```

Mostrar la matriz de correlación con un mapa de calor

```
plt.figure(figsize=(10, 8))
```

```
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
            linewidths=0.5)
```

```
plt.title("Mapa de calor de la correlación entre características")
```

```
plt.show()
```

```
danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
$ python app.py
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

8. Identificación de patrones

Gráfico de caja para comparar la longitud del sépalo entre las especies

```
plt.figure(figsize=(8, 6))
```

```
sns.boxplot(x='species', y='sepal_length', data=data)
```

```
plt.title("Comparación de la longitud del sépalo por especie")
```

```
plt.show()
```

```
danie@LAPTOP-0MT3MCT1 MINGW64 ~/Downloads/Coquito-web-1
$ python app.py
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000