Python & R

Zakaria Rekhla

Introduction

Le code suivant analyse le jeu de données Iris à l'aide de Python. Il calcule des statistiques descriptives telles que la moyenne, la médiane, la variance et la corrélation. De plus, des visualisations telles que des pairplots, des cartes thermiques, des boxplots et des histogrammes sont créées.

Python Code

1- Dataset Loading and Information

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

# Charger le jeu de donn es Iris

from sklearn.datasets import load_iris
data = load_iris()

# Cr er un DataFrame pour faciliter la manipulation

iris_df = pd.DataFrame(data.data, columns=data.feature_names)
iris_df['species'] = pd.Categorical.from_codes(data.target, data.target_names)

# Afficher des informations de base sur le jeu de donn es

print("Aper u du jeu de donn es:")
print(iris_df.head())

print("\nInformations sur le jeu de donn es:")
iris_df.info()

print("\nDescription du jeu de donn es:")
print(iris_df.describe())
```

```
Dataset Head:
   sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
0                5.1               3.5                1.4               0.2
1                4.9               3.0                1.4               0.2
2                4.7               3.2                1.3               0.2
3                4.6               3.1                1.5               0.2
4                5.0               3.6                1.4               0.2

  species
0  setosa
1  setosa
2  setosa
3  setosa
4  setosa

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   sepal length (cm)  150 non-null    float64
 1   sepal width (cm)   150 non-null    float64
 2   petal length (cm)  150 non-null    float64
 3   petal width (cm)   150 non-null    float64
 4   species            150 non-null    category
dtypes: category(1), float64(4)
memory usage: 5.1 KB

Dataset Description:
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.057333           3.758000
std             0.828066          0.435866           1.765298
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)
count        150.000000
mean           1.199333
std            0.762238
min            0.100000
25%            0.300000
50%            1.300000
75%            1.800000
max            2.500000
```

FIGURE 1 – Screenshot of dataset loading and information display.

1- Descriptive Statistics

```python

# Statistiques descriptives

print("\nMoyenne :")
print(iris_df.mean(numeric_only=True))

print("\nM diane :")
print(iris_df.median(numeric_only=True))

print("\nEffectif :")
print(iris_df.count())

print("\nVariance :")
print(iris_df.var(numeric_only=True))

print("\n cart-type :")
print(iris_df.std(numeric_only=True))

```

```
print("\nCovariance :")
print(iris_df.cov())

print("\nCorr lation :")
print(iris_df.corr())
```

```
Moyenne (Mean):
sepal length (cm)    5.843333
sepal width (cm)     3.057333
petal length (cm)    3.758000
petal width (cm)     1.199333
dtype: float64

Médiane (Median):
sepal length (cm)    5.80
sepal width (cm)     3.00
petal length (cm)    4.35
petal width (cm)     1.30
dtype: float64

Effectif (Count):
sepal length (cm)    150
sepal width (cm)     150
petal length (cm)    150
petal width (cm)     150
species              150
dtype: int64

Variance:
sepal length (cm)    0.685694
sepal width (cm)     0.189979
petal length (cm)    3.116278
petal width (cm)     0.581006
dtype: float64

Écart Type (Standard Deviation):
sepal length (cm)    0.828066
sepal width (cm)     0.435866
petal length (cm)    1.765298
petal width (cm)     0.762238
dtype: float64
```

FIGURE 2 – Screenshot of descriptive statistics.

```
Covariance:
                   sepal length (cm)  sepal width (cm)  petal length (cm)  \
sepal length (cm)           0.685694         -0.042434           1.274315
sepal width (cm)           -0.042434          0.189979          -0.329656
petal length (cm)           1.274315         -0.329656           3.116278
petal width (cm)            0.516271         -0.121639           1.295609

                   petal width (cm)
sepal length (cm)          0.516271
sepal width (cm)          -0.121639
petal length (cm)          1.295609
petal width (cm)           0.581006

Corrélation (Correlation):
                   sepal length (cm)  sepal width (cm)  petal length (cm)  \
sepal length (cm)           1.000000         -0.117570           0.871754
sepal width (cm)           -0.117570          1.000000          -0.428440
petal length (cm)           0.871754         -0.428440           1.000000
petal width (cm)            0.817941         -0.366126           0.962865

                   petal width (cm)
sepal length (cm)          0.817941
sepal width (cm)          -0.366126
petal length (cm)          0.962865
petal width (cm)           1.000000
```

FIGURE 3 – Screenshot of descriptive statistics.

3- Visualizations

```
1
2  # Visualisations
3
4  sns.pairplot(iris_df, hue='species', corner=True)
5  plt.suptitle('Pairplot du Jeu de Donn es Iris', y=1.02)
6  plt.show()
7
8  # Carte thermique des corr lations
9
10 plt.figure(figsize=(10, 8))
11 sns.heatmap(iris_df.corr(), annot=True, fmt='.2f', cmap='coolwarm')
12 plt.title('Carte Thermique des Corr lations')
13 plt.show()
14
15 # Boxplot pour chaque caract ristique selon les esp ces
16
17 plt.figure(figsize=(15, 10))
18 iris_features = iris_df.columns[:-1]
19 for i, feature in enumerate(iris_features, 1):
20     plt.subplot(2, 2, i)
21     sns.boxplot(data=iris_df, x='species', y=feature)
22     plt.title(f'Boxplot de {feature}')
23 plt.tight_layout()
24 plt.show()
25
26 # Histogrammes pour chaque caract ristique
27
28 iris_df.hist(bins=15, figsize=(12, 10), edgecolor='black')
29 plt.suptitle('Histogrammes des Caract ristiques d\'Iris')
30 plt.show()
```

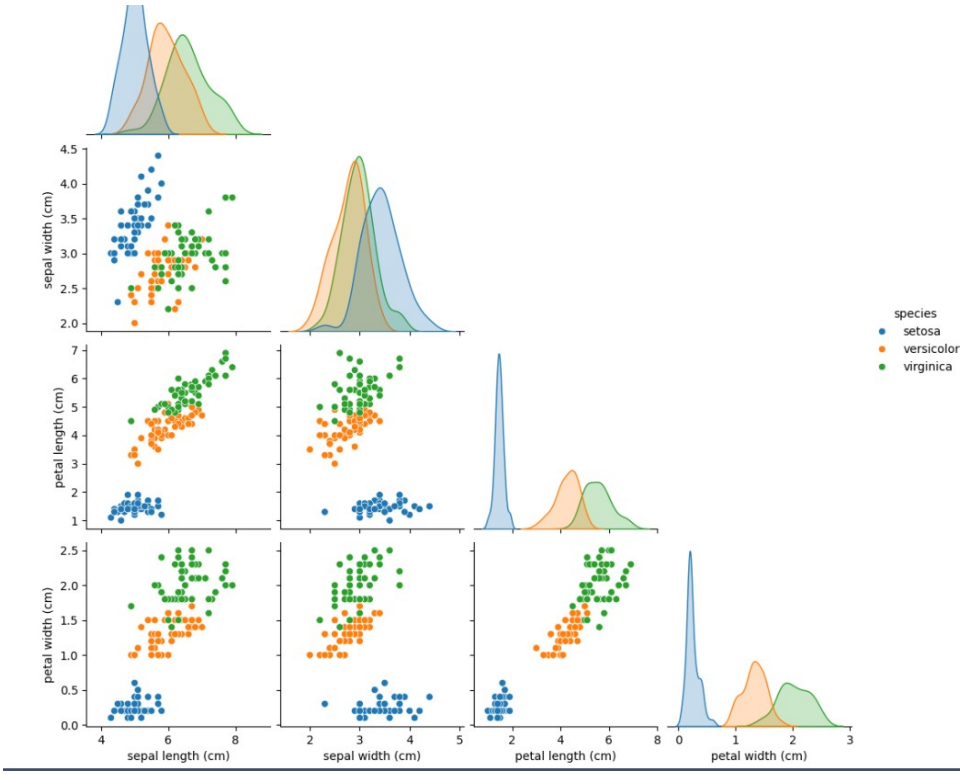

FIGURE 4 – Pairplot du Jeu de Données Iris.

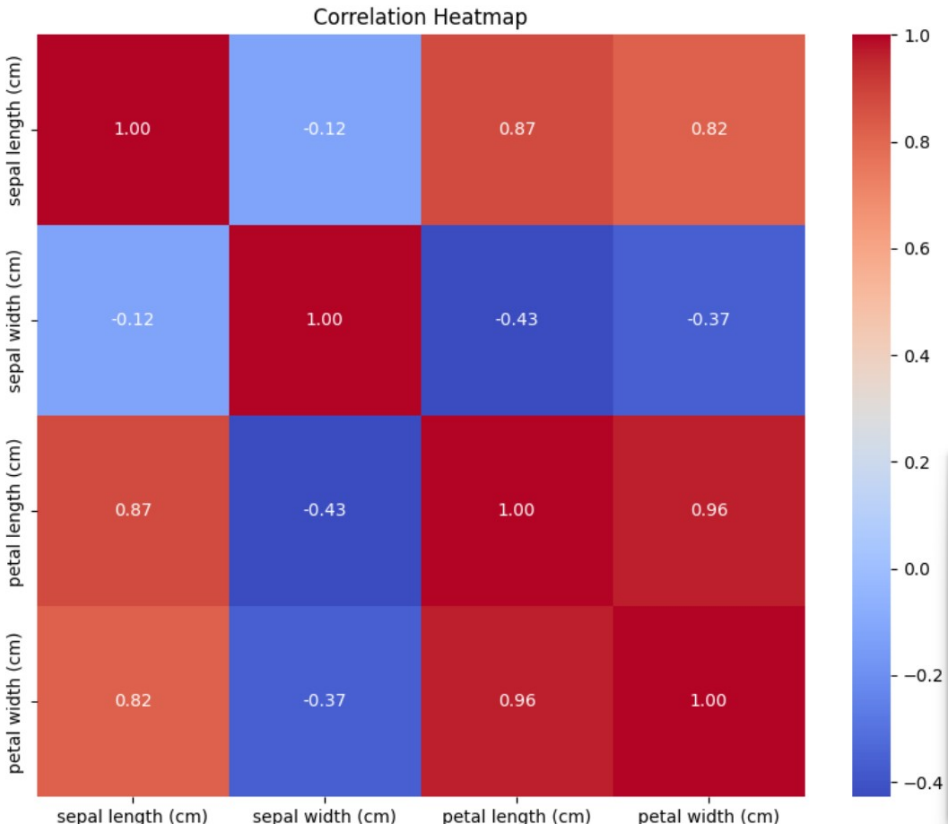

FIGURE 5 – Carte Thermique des Corrélations.

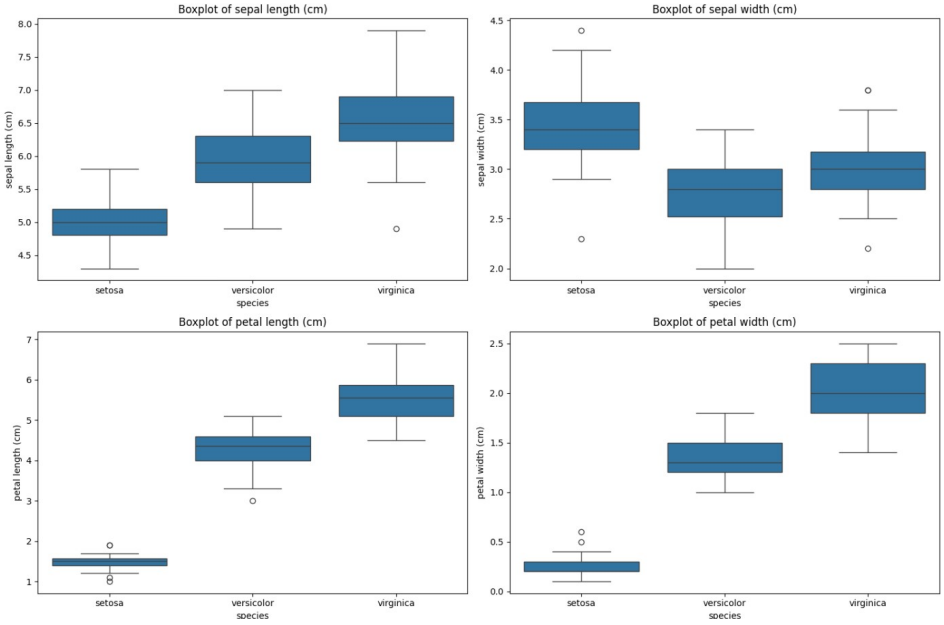

FIGURE 6 – Boxplot .

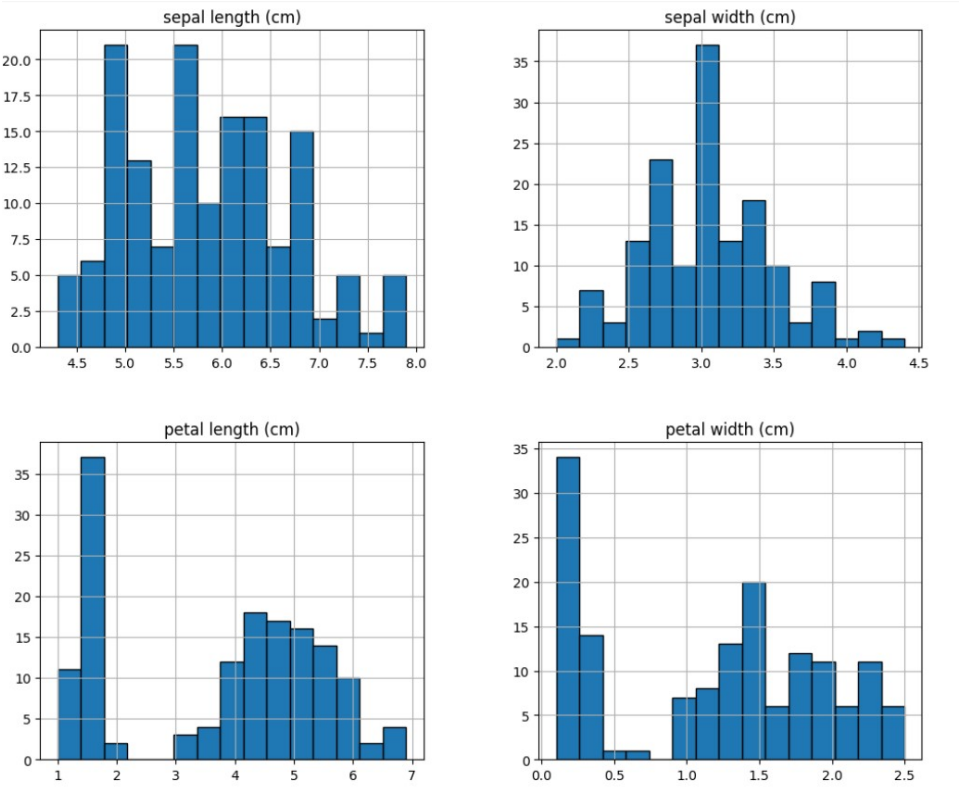

FIGURE 7 – Histogrammes des Caractéristiques dÍris .