Name:Ankita Dhokane
Roll No: 18
Class: SYMCA-B
Subject: Data Science

Assignment No: 5 - 1.Write a Python script for Correlation and Covariance
a. Find the correlation matrix.
b. Plot the correlation plot on dataset and visualize giving an overview of relationships among any dataset.

CODE:

```
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load or create a dataset (for demonstration purposes, using random data)
# Replace this with your dataset
np.random.seed(42)
data = pd.DataFrame({
    'A': np.random.randn(100),
    'B': np.random.randn(100),
    'C': np.random.randn(100) + 0.5,
    'D': 2.5 * np.random.randn(100) - 1
})

# Display the first few rows of the dataset
print("Dataset:")
print(data.head())
```

```
# a. Find the correlation matrix
correlation_matrix = data.corr()
print("\nCorrelation Matrix:")
print(correlation_matrix)

# b. Find the covariance matrix
covariance_matrix = data.cov()
print("\nCovariance Matrix:")
print(covariance_matrix)

# Plotting the correlation matrix using a heatmap
plt.figure(figsize=(8, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0)
plt.title('Correlation Heatmap')
plt.show()
```

OUTPUT:

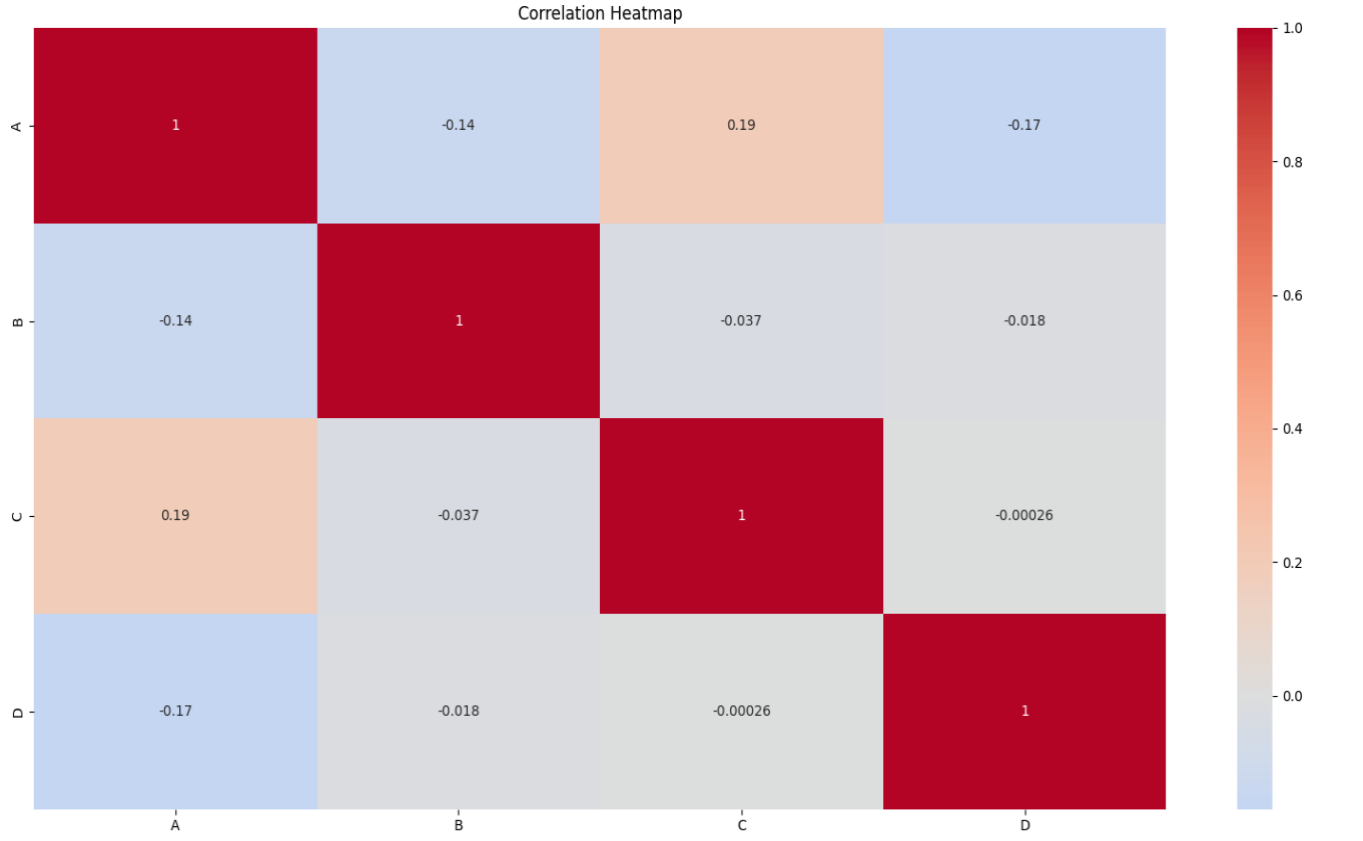
IDLE Shell 3.12.4

File Edit Shell Debug Options Window Help

```
Python 3.12.4 (tags/v3.12.4:8e8a4ba, Jun  6 2024, 19:30:16) [MSC v.1940 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: C:/Users/ADMIN/AppData/Local/Programs/Python/Python312/correlationcovariance.py
Dataset:
          A         B         C         D
0  0.496714 -1.415371  0.857787 -3.072488
1 -0.138264 -0.420645  1.060785 -2.400453
2  0.647689 -0.342715  1.583051  0.868234
3  1.523030 -0.802277  1.553802  0.525926
4 -0.234153 -0.161286 -0.877669 -1.052254

Correlation Matrix:
          A         B         C         D
A  1.000000 -0.136422  0.190840 -0.170227
B -0.136422  1.000000 -0.036632 -0.017613
C  0.190840 -0.036632  1.000000 -0.000259
D -0.170227 -0.017613 -0.000259  1.000000

Covariance Matrix:
          A         B         C         D
A  0.824770 -0.118154  0.187922 -0.341693
B -0.118154  0.909484 -0.037879 -0.037125
C  0.187922 -0.037879  1.175669 -0.000621
D -0.341693 -0.037125 -0.000621  4.885221
```


Correlation Heatmap
A
B
C
D
1
-0.14
0.19
-0.17
-0.14
1
-0.037
-0.018
0.19
-0.037
1
-0.00026
-0.17
-0.018
-0.00026
1
A
B
C
D
1.0
0.8
0.6
0.4
0.2
0.0