PRACTICAL NO 4

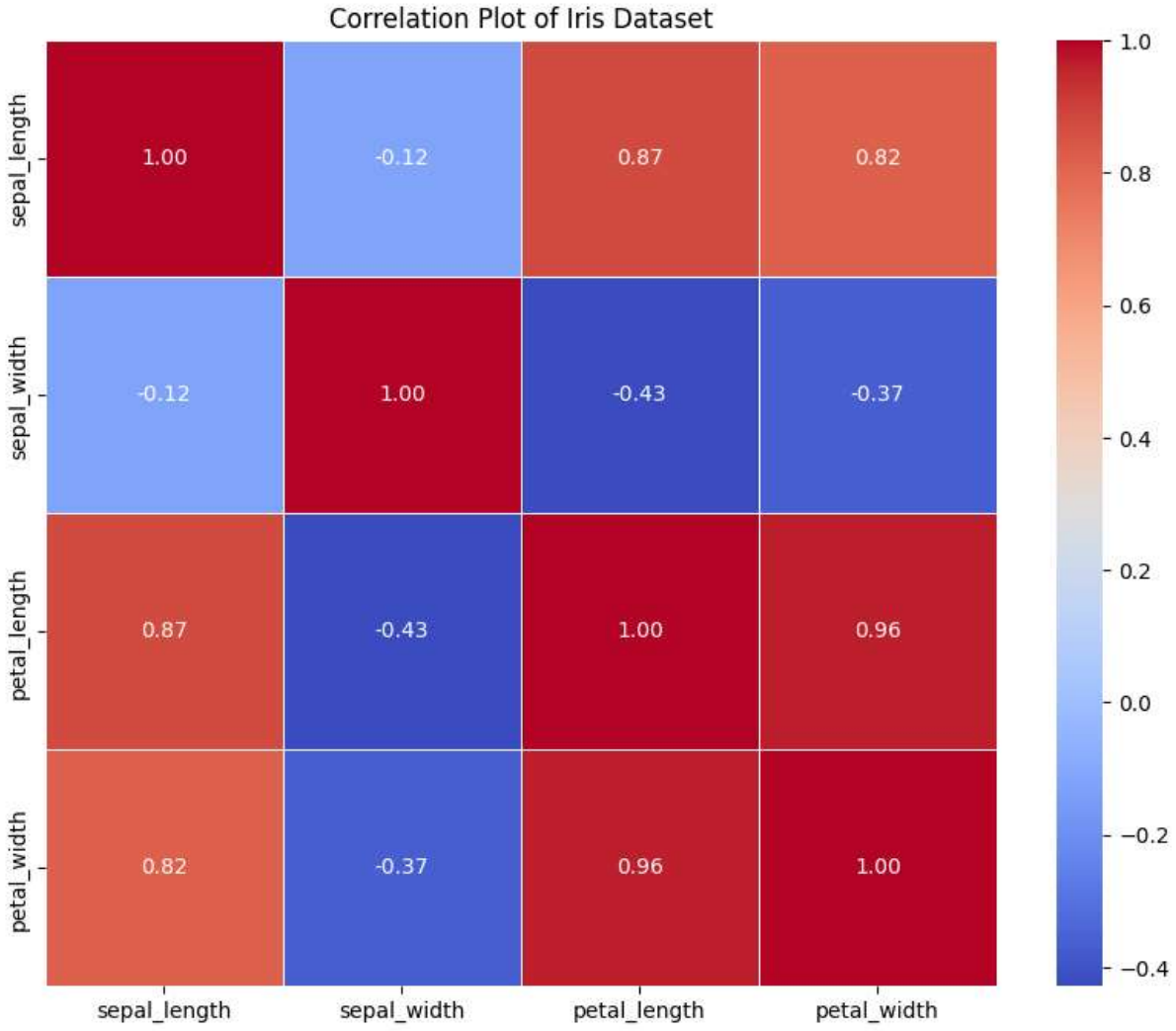
Code :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import f_oneway

# Load the Iris dataset from Seaborn
iris_data = sns.load_dataset('iris')

# a. Find the correlation matrix:
# Exclude the 'species' column for correlation calculation
correlation_matrix = iris_data.drop('species', axis=1).corr()
print("Correlation Matrix:")
print(correlation_matrix)
print("\n")

# b. Plot the correlation plot on the dataset:
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", fmt=".2f",
linewidths=.5)
plt.title("Correlation Plot of Iris Dataset")
plt.show()

# c. Analysis of Covariance (ANOVA) if data have categorical variables:
# Assuming 'species' is a categorical variable (replace with your actual
column name)
categories = iris_data['species'].unique()

# Perform ANOVA for each feature based on the categorical variable
for feature in iris_data.columns[:-1]:  # Exclude the last column
(assuming it's the categorical variable)
    groups = [iris_data[iris_data['species'] == category][feature] for
category in categories]
    f_statistic, p_value = f_oneway(*groups)

    print(f"ANOVA for {feature}: F-statistic = {f_statistic}, p-value =
{p_value}")
```

Output :

Correlation Matrix:

	sepal_length	sepal_width	petal_length	petal_width
sepal_length	**1.000000**	**-0.117570**	**0.871754**	**0.817941**
sepal_width	**-0.117570**	**1.000000**	**-0.428440**	**-0.366126**
petal_length	**0.871754**	**-0.428440**	**1.000000**	**0.962865**
petal_width	**0.817941**	**-0.366126**	**0.962865**	**1.000000**



ANOVA for sepal_length: F-statistic = 119.26450218450468, p-value = 1.6696691907693826e-31

ANOVA for sepal_width: F-statistic = 49.160040089612075, p-value = 4.492017133309115e-17

ANOVA for petal_length: F-statistic = 1180.161182252981, p-value = 2.8567766109615584e-91

ANOVA for petal_width: F-statistic = 960.007146801809, p-value = 4.169445839443116e-85