

National Forensic Sciences University
School of Cyber Security and Digital Forensics
Assignment 3 : Matplot + Pandas, Artificial Intelligence

NAME - NEERAJ KUMAR

1. Load Iris Dataset.

```
import pandas as pd

from sklearn.datasets import load_iris

iris = load_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature_names)

df['target'] = iris.target

df['target_names'] = df['target'].map(lambda x:iris.target_names[x])

print(df.head())
```

OUTPUT

```
➡  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

   target target_names
0       0      setosa
1       0      setosa
2       0      setosa
3       0      setosa
4       0      setosa
```

2. Draw a line Graph for Sepal length vs Sepal width.

```
import matplotlib.pyplot as plt

import pandas as pd

from sklearn.datasets import load_iris

iris = load_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature_names)

sepal_length = df['sepal length (cm)']

sepal_width = df['sepal width (cm)']

plt.plot(sepal_length, sepal_width, marker='o')

plt.title('Sepal Length vs. Sepal Width')

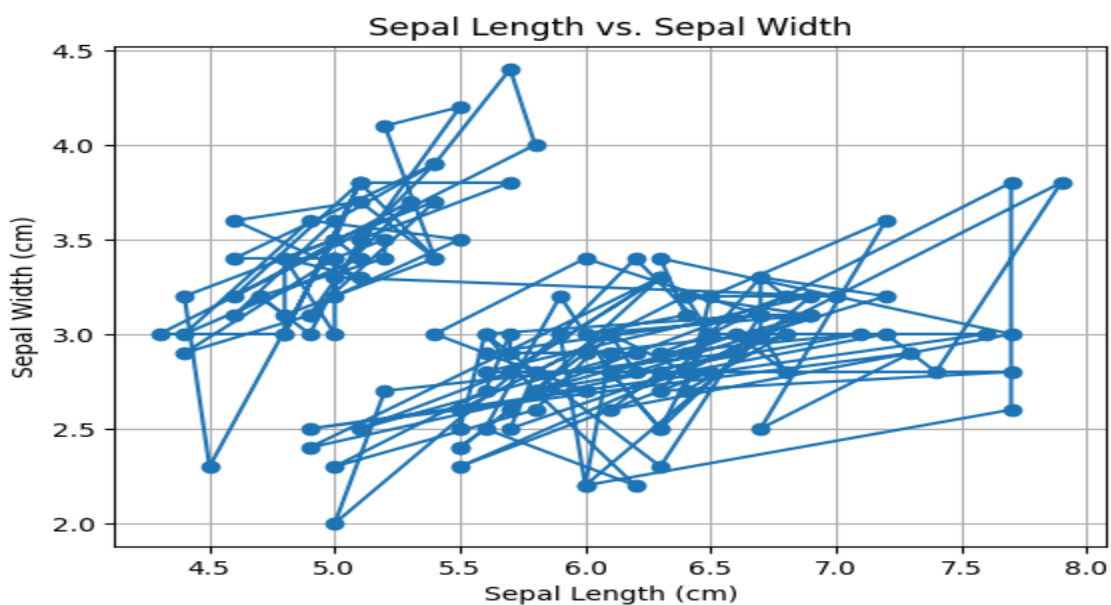
plt.xlabel('Sepal Length (cm)')

plt.ylabel('Sepal Width (cm)')

plt.grid(True)

plt.show()
```

OUTPUT



3. Draw a Bar Graph for Petal Length vs Petal Width.

```
import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from sklearn.datasets import load_iris

iris = load_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature_names)

petal_length = df['petal length (cm)']

petal_width = df['petal width (cm)']

plt.bar(petal_length, petal_width, color='skyblue')

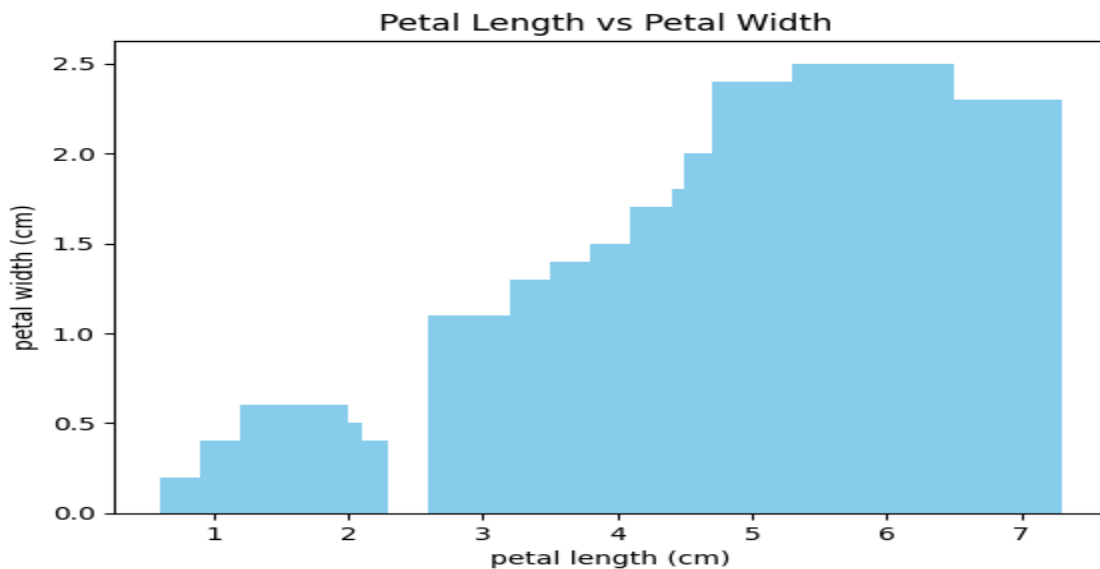
plt.xlabel('petal length (cm)')

plt.ylabel('petal width (cm)')

plt.title(' Petal Length vs Petal Width  ')

plt.show()
```

OUTPUT



4. Draw a Histogram Graph for Sepal Length

```
import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

from sklearn.datasets import load_iris

iris = load_iris()

df = pd.DataFrame(data=iris.data, columns=iris.feature_names)

sepal_length = df['sepal length (cm)']

plt.hist(sepal_length, color='skyblue')

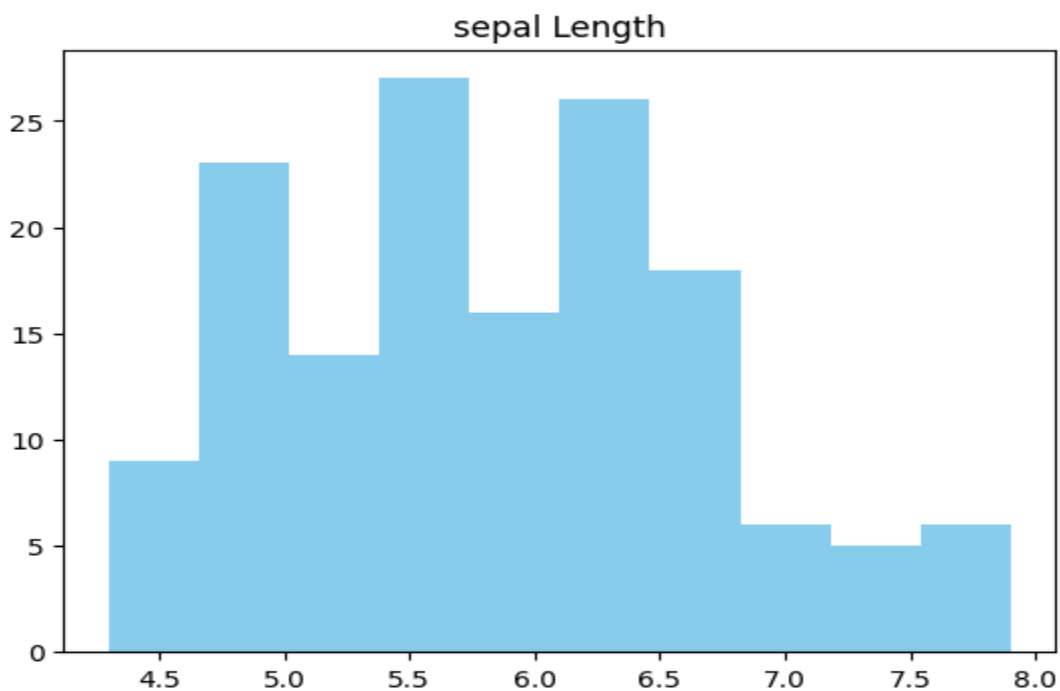
# plt.xlabel('petal length (cm)')

# plt.ylabel('petal width (cm)')

plt.title('sepal Length')

plt.show()
```

OUTPUT



Using the Pandas Library write the Python script for the following:

1. Read all CSV files in a given folder and Display File Name as well as File Content (Hint: glob)

```
import os

import pandas as pd

folder_path = r'C:\Users\CBI\Desktop\COLLAB\matplotlib'

for filename in os.listdir(folder_path):

    if filename.endswith('.csv'):

        file_path = os.path.join(folder_path, filename)

        df = pd.read_csv(file_path)

        print(f"the filename that is being printed is: '{filename}")

        print('\n')

        print(f"Content:")

        print(df)

        print('\n')
```

OUTPUT

```
the filename that is being printed is: 'file1.csv
Content:
   Name  Age
0  Alice   25
1   Bob   30
the filename that is being printed is: 'file2.csv
Content:
   Name  Age
0  Charlie 35
1   David 40
the filename that is being printed is: 'file3.csv
Content:
   Name  Age
0   Eve   28
1  Frank 33
the filename that is being printed is: 'file4.csv
Content:
   Name  Age
0  Grace 29
1   Hank 32
the filename that is being printed is: 'file5.csv
Content:
   Name  Age
0   Ivy  26
1  Jack  34
...
Content:
   Name  Age
0   Liam 31
1    Mia 27
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

2. Read multiple Excel files (file.xlsx, file3.xlsx, and file2.xlsx) containing multiple sheets

```
import pandas as pd

import os

directory_path = r'C:\Users\CBI\Desktop\COLLAB\matplotlib'

for filename in os.listdir(directory_path):

    if filename.endswith('.xlsx'):

        file_path = os.path.join(directory_path, filename)

        dfs = pd.read_excel(file_path, sheet_name=None)

        print(f"Current Excel File: {filename}")

        for sheet_name, df in dfs.items():

            print(f'Sheet Name: {sheet_name}')

            print(df)

            print('-' * 40)
```

OUTPUT

```
... Current Excel File: sample_large_data - Copy (2).xlsx
Sheet Name: Sheet1
  Unnamed: 0  Unnamed: 1  Unnamed: 2
0      sdhthft      NaN      NaN
1      NaN      tfh      NaN
2      NaN      NaN      NaN
3      NaN      NaN      NaN
4      zftjx      NaN      NaN
5      NaN      NaN      NaN
6      jxf      NaN      NaN
7      NaN      NaN      NaN
8      NaN      NaN      mxfyjxf
9      NaN      NaN      NaN
10     NaN      NaN      NaN
11     NaN      NaN      NaN
12     NaN      NaN      NaN
13     NaN      zftjxftj      xfdt
-----
Sheet Name: Sheet3
Empty DataFrame
Columns: [efesfsegetjteregsg]
Index: []
-----
Sheet Name: Sheet2
Empty DataFrame
...
Empty DataFrame
Columns: [this is a sheet 2 content in excel]
Index: []
-----
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...
```

3. Read Multiple Sheets and Print a Specific One By Its Name.

```
import pandas as pd

import os

file_path =
r'C:\Users\CBI\Desktop\COLLAB\matplotlib\sample_large_data.xlsx'

dfs = pd.read_excel(file_path, sheet_name=None)

specific_sheet_one = 'Sheet1'

if specific_sheet_one in dfs:

    print(f"the content of {specific_sheet_one}:")

    print(dfs[specific_sheet_one])

else:

    print(f"{specific_sheet_one} was not found in the Excel file.")
```

OUTPUT

```
the content of Sheet1:
Empty DataFrame
Columns: [this is a excel sheet page 1`]
Index: []
```