

PRACTICAL - 05

5. Taking Iris data, plot the following with proper legend and axis labels:
(Download IRIS data from: <https://archive.ics.uci.edu/ml/datasets/iris>
(<https://archive.ics.uci.edu/ml/datasets/iris>) or import it from sklearn.datasets)

- Plot bar chart to show the frequency of each class label in the data.
- Draw a scatter plot for Petal width vs sepal width.
- Plot density distribution for feature petal length.
- Use a pair plot to show pairwise bivariate distribution in the Iris Dataset.

CODE :

```
[90] import numpy as np
import pandas as pd
import sklearn.datasets as sns
import matplotlib.pyplot as plt
import seaborn as sns

[73] iris=sns.load_iris()

[74] print(iris.DESCR)
```

```
.. _iris_dataset:
```

```
Iris plants dataset  
-----
```

```
**Data Set Characteristics:**
```

```
:Number of Instances: 150 (50 in each of three classes)  
:Number of Attributes: 4 numeric, predictive attributes and the class  
:Attribute Information:  
  - sepal length in cm  
  - sepal width in cm  
  - petal length in cm  
  - petal width in cm  
  - class:  
    - Iris-Setosa  
    - Iris-Versicolour  
    - Iris-Virginica
```

```
:Summary Statistics:
```

```
=====  Min  Max   Mean   SD   Class Correlation  
=====  ----  ----  -----  ----  -----  
sepal length:  4.3  7.9   5.84   0.83   0.7826  
sepal width:   2.0  4.4   3.05   0.43  -0.4194  
petal length:   1.0  6.9   3.76   1.76   0.9490 (high!)  
petal width:   0.1  2.5   1.20   0.76   0.9565 (high!)  
=====
```

```
:Missing Attribute Values: None  
:Class Distribution: 33.3% for each of 3 classes.  
:Creator: R.A. Fisher  
:Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)  
:Date: July, 1988
```

The famous Iris database, first used by Sir R.A. Fisher. The dataset is taken from Fisher's paper. Note that it's the same as in R, but not as in the UCI Machine Learning Repository, which has two wrong data points.

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2; the latter are NOT linearly separable from each other.

```
.. topic:: References
```

- Fisher, R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).
- Duda, R.O., & Hart, P.E. (1973) Pattern Classification and Scene Analysis. (Q327.D83) John Wiley & Sons. ISBN 0-471-22361-1. See page 218.
- Dasarthy, B.V. (1980) "Nosing Around the Neighborhood: A New System Structure and Classification Rule for Recognition in Partially Exposed Environments". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-2, No. 1, 67-71.
- Gates, G.W. (1972) "The Reduced Nearest Neighbor Rule". IEEE Transactions on Information Theory, May 1972, 431-433.
- See also: 1988 MLC Proceedings, 54-64. Cheeseman et al's AUTOCLASS II conceptual clustering system finds 3 classes in the data.
- Many, many more ...

```
iris.feature_names
```

[75]

```
... ['sepal length (cm)',  
      'sepal width (cm)',  
      'petal length (cm)',  
      'petal width (cm)']
```

```
iris.target
```

[76]

```
... array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
          0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
          0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
          1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,  
          1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,  
          2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,  
          2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2])
```

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

[77]

```
df.head()
```

[78]

	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

```
df['Iris type'] = iris['target']
```

[79]

```
df
```

[80]

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	Iris type
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0
...
145	6.7	3.0	5.2	2.3	2
146	6.3	2.5	5.0	1.9	2
147	6.5	3.0	5.2	2.0	2
148	6.2	3.4	5.4	2.3	2
149	5.9	3.0	5.1	1.8	2

150 rows × 5 columns

```
[81] def convert(x):  
    if x == 0:  
        val = 'setosa'  
    elif x == 1:  
        val = 'versicolor'  
    else:  
        val = 'virginica'  
    return val
```

```
[86] df['type name']=df['Iris type'].apply(convert)
```

```
[87] df.head()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	Iris type	type name
0	5.1	3.5	1.4	0.2	0	setosa
1	4.9	3.0	1.4	0.2	0	setosa
2	4.7	3.2	1.3	0.2	0	setosa
3	4.6	3.1	1.5	0.2	0	setosa
4	5.0	3.6	1.4	0.2	0	setosa

a. Plot bar chart to show the frequency of each class label in the data.

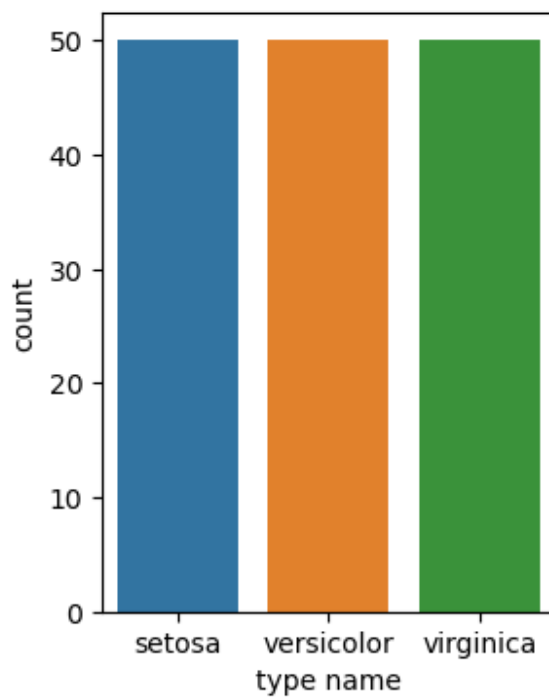
```
[89] df.groupby('type name').count()
```

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	Iris type
type name					
setosa	50	50	50	50	50
versicolor	50	50	50	50	50
virginica	50	50	50	50	50

```
[97] plt.figure(figsize=(3,4))  
sns.countplot(x='type name',data=df)
```

```
... <Axes: xlabel='type name', ylabel='count'>
```

...



b. Draw a scatter plot for Petal width vs sepal width.

```
df.columns
```

[98]

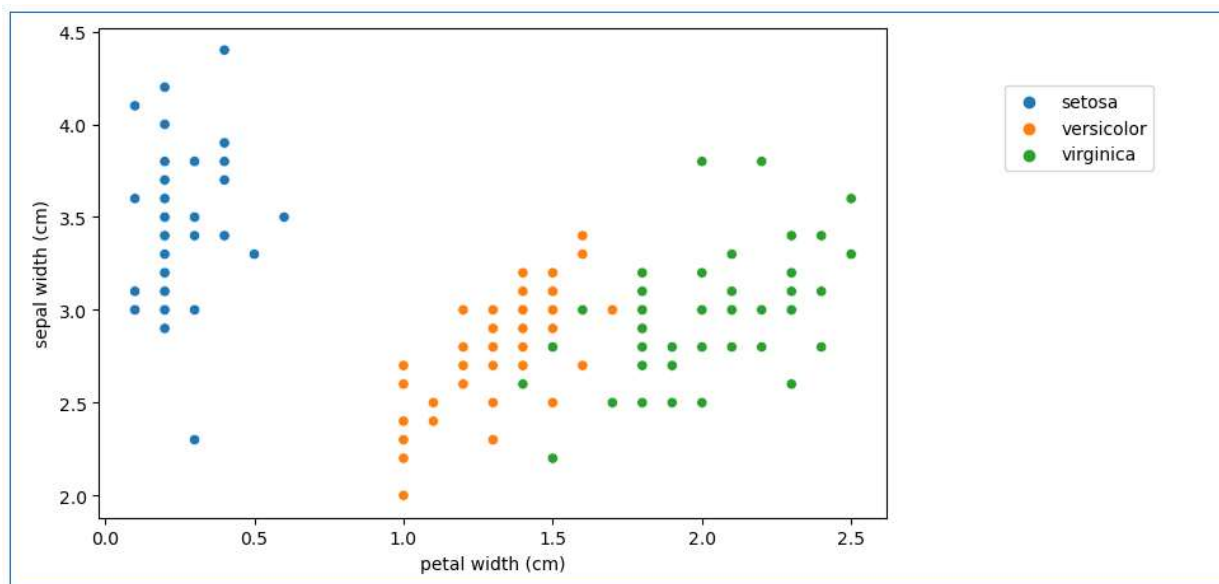
```
... Index(['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)',  
         'petal width (cm)', 'Iris type', 'type name'],  
        dtype='object')
```

▷

```
plt.figure(figsize=(8,5))  
sns.scatterplot(x='petal width (cm)',y='sepal width (cm)',data=df,hue='type name')  
plt.legend(bbox_to_anchor=(1.35,0.9))
```

[108]

```
... <matplotlib.legend.Legend at 0x12c41ab90>
```



SUBMITTED TO : DR. PRITI JAGWANI

NAME : GARIMA ROHILLA

ROLL NO : CSC/21/37

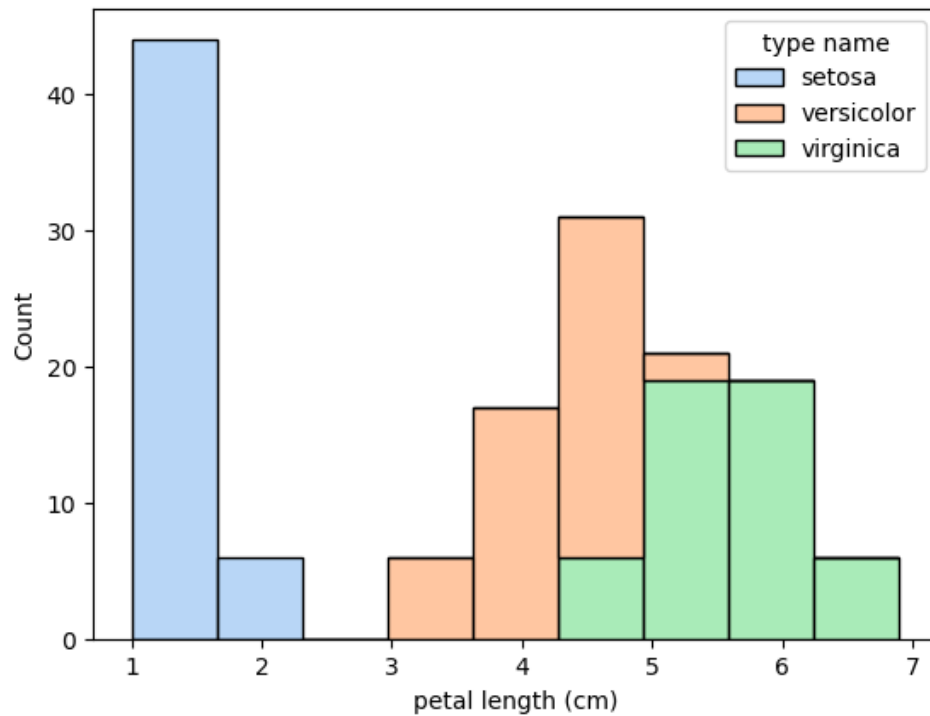
UNIVERSITY ROLL NO : 21059570010

c. Plot density distribution for feature petal length.

```

> sns.histplot(x='petal length (cm)',data=df,hue='type name',multiple='stack',palette='pastel')
[127]
... <Axes: xlabel='petal length (cm)', ylabel='Count'>

```



d. Use a pair plot to show pairwise bivariate distribution in the Iris Dataset.

