practical3

April 15, 2025

[33]:
```
import pandas as pd
```

[34]:
```
import numpy as np
```

[35]:
```
import matplotlib.pyplot as plt
```

[36]:
```
import seaborn as sns
```

[37]:
```
ssd = pd.read_csv("Iris_Dataset.csv")
```

[38]:
```
ssd3 = pd.read_csv("iris.csv")
```

[39]:
```
ssd3
```

[39]:
```
     150    4  setosa  versicolor  virginica
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 x 5 columns]
```

[40]:
```
ssd
```

[40]:
```
     sepal_length  sepal_width  petal_length  petal_width         species
0             5.1          3.5           1.4          0.2          setosa
1             4.9          3.0           1.4          0.2          setosa
2             4.7          3.2           1.3          0.2          setosa
3             4.6          3.1           1.5          0.2          setosa
4             5.0          3.6           1.4          0.2          setosa
```

```
..            …           …           …          …              …
145         6.7         3.0         5.2        2.3  Iris-virginica
146         6.3         2.5         5.0        1.9  Iris-virginica
147         6.5         3.0         5.2        2.0  Iris-virginica
148         6.2         3.4         5.4        2.3  Iris-virginica
149         5.9         3.0         5.1        1.8  Iris-virginica

[150 rows x 5 columns]
```

[41]:
```
ssd.isnull().sum()
```

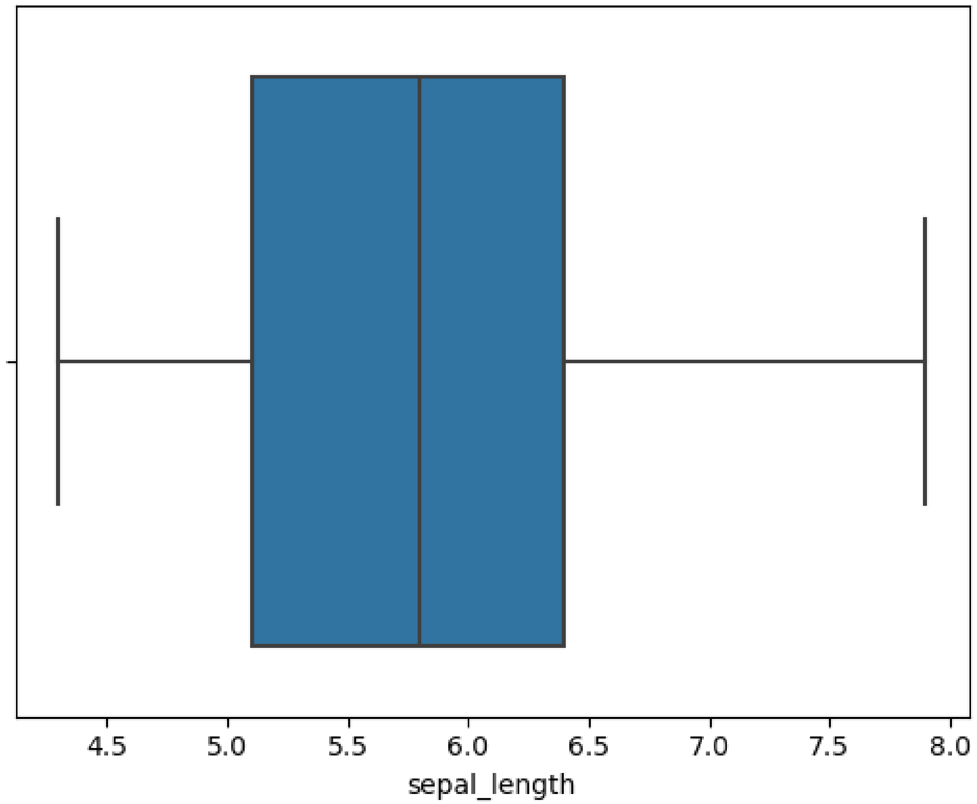
[41]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

[42]:
```
ssd.describe(include="all")
```

[42]:
```
       sepal_length  sepal_width  petal_length  petal_width species
count    150.000000   150.000000    150.000000   150.000000     150
unique          NaN          NaN           NaN          NaN       3
top             NaN          NaN           NaN          NaN  setosa
freq            NaN          NaN           NaN          NaN      50
mean       5.843333     3.054000      3.758667     1.198667     NaN
std        0.828066     0.433594      1.764420     0.763161     NaN
min        4.300000     2.000000      1.000000     0.100000     NaN
25%        5.100000     2.800000      1.600000     0.300000     NaN
50%        5.800000     3.000000      4.350000     1.300000     NaN
75%        6.400000     3.300000      5.100000     1.800000     NaN
max        7.900000     4.400000      6.900000     2.500000     NaN
```

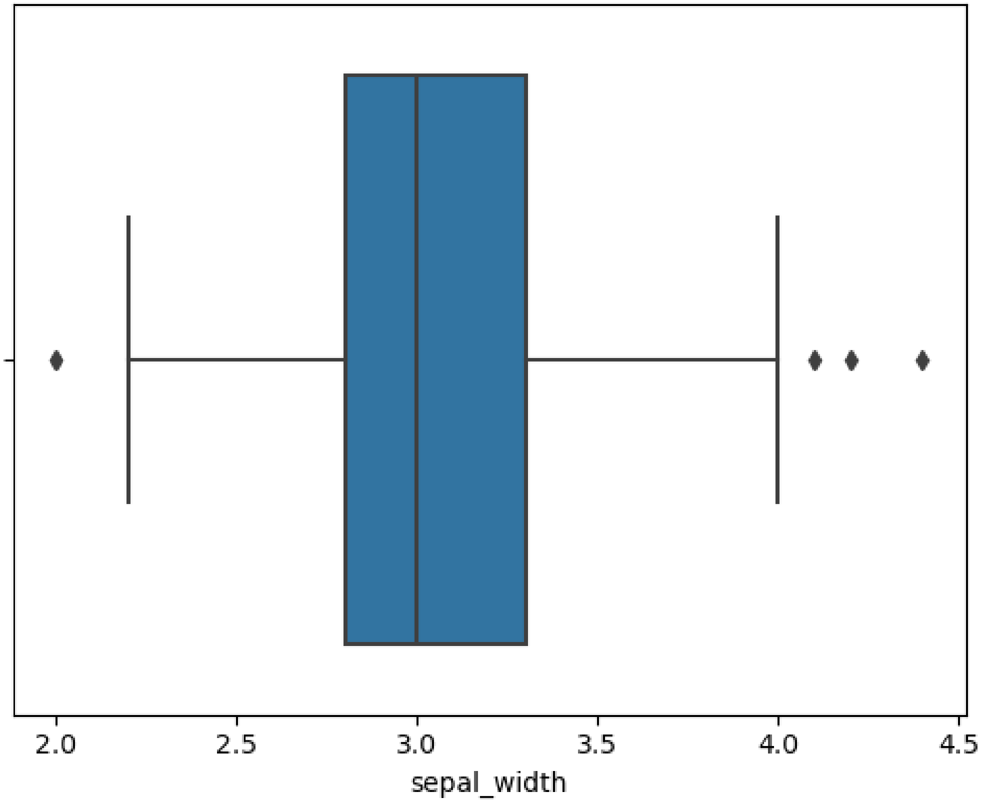
[43]:
```
sns.boxplot(x="sepal_length",data=ssd)
```

[43]:
```
<Axes: xlabel='sepal_length'>
```



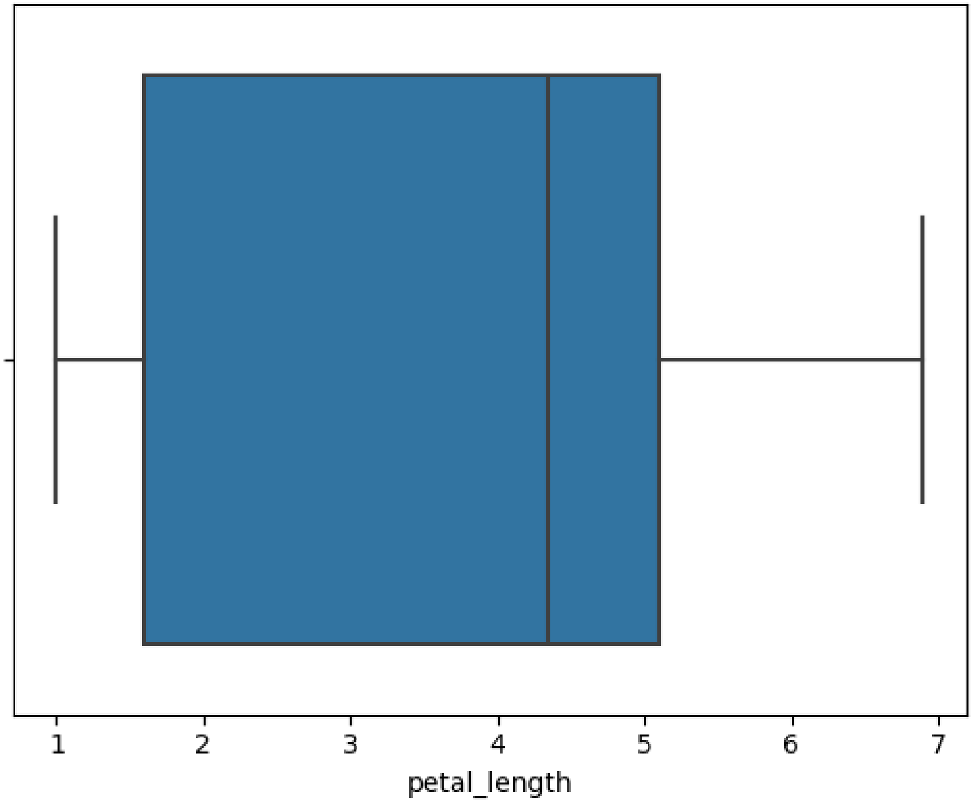
```
[45]: sns.boxplot(x="sepal_width",data=ssd)
```

```
[45]: <Axes: xlabel='sepal_width'>
```



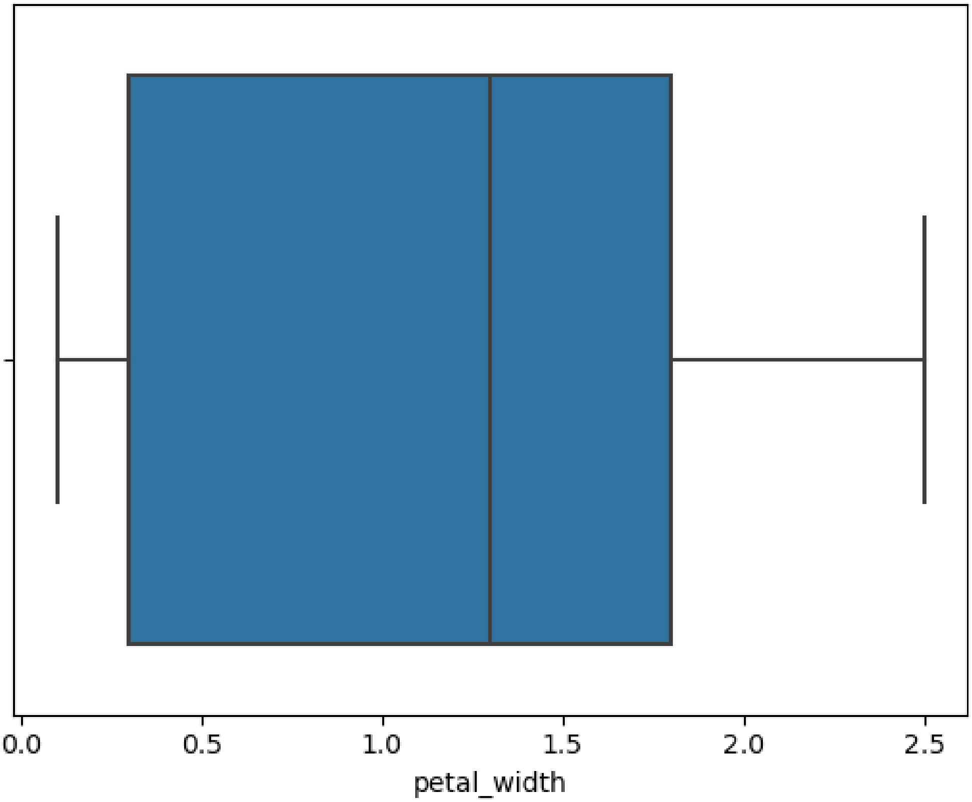
```
[46]: sns.boxplot(x="petal_length",data=ssd)
```

```
[46]: <Axes: xlabel='petal_length'>
```



```
[47]: sns.boxplot(x="petal_width",data=ssd)
```

```
[47]: <Axes: xlabel='petal_width'>
```



[30]:
```
ssd.species.unique()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[30], line 1
----> 1 ssd.species.unique()

AttributeError: 'int' object has no attribute 'species'
```

[31]:
```
from sklearn.preprocessing import LabelEncoder
```

[27]:
```
le = LabelEncoder()
```

[28]:
```
le.fit_transform(ssd["species"])
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[28], line 1
----> 1 le.fit_transform(ssd["species"])
```

```
TypeError: 'int' object is not subscriptable
```

```
[29]: ssd["species"]=le.fit_transform(ssd["species"])
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[29], line 1
----> 1 ssd["species"]=le.fit_transform(ssd["species"])

TypeError: 'int' object is not subscriptable
```

```
[25]: ssd
```

```
[25]: 0
```

```
[24]: ssd.species.unique()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[24], line 1
----> 1 ssd.species.unique()

AttributeError: 'int' object has no attribute 'species'
```

```
[23]: g=ssd.groupby("species")
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Cell In[23], line 1
----> 1 g=ssd.groupby("species")

AttributeError: 'int' object has no attribute 'groupby'
```

```
[22]: for ssd,species_vvk in g:
          print(vvk)
          print(".....................")
          print(species_vvk)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[22], line 2
      1 for ssd,species_vvk in g:
----> 2     print(vvk)
      3     print("…")
      4     print(species_vvk)

NameError: name 'vvk' is not defined
```

[40]:

```
g.agg({'sepal_length':'mean'})
```

[40]:

```
         sepal_length
species
0               5.936
1               6.588
2               5.006
```

[41]:

```
g.agg({"sepal_length":"mean","sepal_width":"mean","petal_length":
 ↪"mean","petal_width":"mean"})
```

[41]:

```
         sepal_length  sepal_width  petal_length  petal_width
species
0               5.936        2.770         4.260        1.326
1               6.588        2.974         5.552        2.026
2               5.006        3.418         1.464        0.244
```

[42]:

```
print ("Median of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": "mean", "sepal_width": "mean", "petal_length":
 ↪"mean","petal_width":"mean"})
```

```
Median of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

[42]:

```
         sepal_length  sepal_width  petal_length  petal_width
species
0               5.936        2.770         4.260        1.326
1               6.588        2.974         5.552        2.026
2               5.006        3.418         1.464        0.244
```

[43]:

```
print ("Median of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": "median", "sepal_width": "median", "petal_length":
 ↪"median","petal_width":"median"})
```

```
Median of Eeach Species
```

```
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

```
[43]:          sepal_length  sepal_width  petal_length  petal_width
       species
       0                 5.9          2.8          4.35          1.3
       1                 6.5          3.0          5.55          2.0
       2                 5.0          3.4          1.50          0.2
```

```
[44]: print ("std of Eeach Species")
      print("0-Iris Setosa")
      print("1-Iris Versicolor")
      print("2-Iris Verginia")
      g.agg ({"sepal_length": "std", "sepal_width": "std", "petal_length":
        ↪"std","petal_width":"std"})
```

```
std of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

```
[44]:          sepal_length  sepal_width  petal_length  petal_width
       species
       0            0.516171     0.313798      0.469911     0.197753
       1            0.635880     0.322497      0.551895     0.274650
       2            0.352490     0.381024      0.173511     0.107210
```

```
[45]: print (" 25 percentile of Eeach Species")
      print("0-Iris Setosa")
      print("1-Iris Versicolor")
      print("2-Iris Verginia")
      g.agg ({"sepal_length": lambda x:x.quantile (0.25), "sepal_width": lambda x:x.
        ↪quantile (0.25),"petal_width": lambda x:x.quantile (0.25),"sepal_length":␣
        ↪lambda x:x.quantile (0.25)})
```

```
 25 percentile of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

```
[45]:          sepal_length  sepal_width  petal_width
       species
       0               5.600        2.525          1.2
       1               6.225        2.800          1.8
       2               4.800        3.125          0.2
```

[46]:

```
print (" 50 percentile of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": lambda x:x.quantile (0.50), "sepal_width": lambda x:x.
 ↪quantile (0.50),"petal_width": lambda x:x.quantile (0.50),"sepal_length":␣
 ↪lambda x:x.quantile (0.50)})
```

```
 50 percentile of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

[46]:

```
         sepal_length  sepal_width  petal_width
species
0                 5.9          2.8          1.3
1                 6.5          3.0          2.0
2                 5.0          3.4          0.2
```

[47]:

```
print (" 75 percentile of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": lambda x:x.quantile (0.75), "sepal_width": lambda x:x.
 ↪quantile (0.75),"petal_width": lambda x:x.quantile (0.75),"sepal_length":␣
 ↪lambda x:x.quantile (0.75)})
```

```
 75 percentile of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

[47]:

```
         sepal_length  sepal_width  petal_width
species
0                 6.3        3.000          1.5
1                 6.9        3.175          2.3
2                 5.2        3.675          0.3
```

[48]:

```
print ("min of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": "min", "sepal_width": "min", "petal_length":
 ↪"min","petal_width":"min"})
```

```
min of Eeach Species
0-Iris Setosa
1-Iris Versicolor
```

```
2-Iris Verginia
```

[48]:
```
         sepal_length  sepal_width  petal_length  petal_width
species
0                 4.9          2.0           3.0          1.0
1                 4.9          2.2           4.5          1.4
2                 4.3          2.3           1.0          0.1
```

[49]:
```
print ("max of Eeach Species")
print("0-Iris Setosa")
print("1-Iris Versicolor")
print("2-Iris Verginia")
g.agg ({"sepal_length": "max", "sepal_width": "max", "petal_length":
 ↪"max","petal_width":"max"})
```

```
max of Eeach Species
0-Iris Setosa
1-Iris Versicolor
2-Iris Verginia
```

[49]:
```
         sepal_length  sepal_width  petal_length  petal_width
species
0                 7.0          3.4           5.1          1.8
1                 7.9          3.8           6.9          2.5
2                 5.8          4.4           1.9          0.6
```

[50]:
```
vvk3.skew(axis = 0, skipna = True)
```

[50]:
```
150           0.314911
4             0.334053
setosa       -0.274464
versicolor   -0.104997
virginica     0.000000
dtype: float64
```

[51]:
```
vvk3.skew(axis = 1, skipna = True)
```

[51]:
```
0      0.652099
1      0.784302
2      0.660432
3      0.579729
4      0.586588
         …
145    0.768184
146    0.786525
147    0.665191
148    0.392508
149    0.463271
```

```
Length: 150, dtype: float64
```

```
[ ]:
```