haridas DSBDA3

March 6, 2025

[14]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[15]:
```
hari=pd.read_csv("iris_dataset.csv")
```

[16]:
```
hari
```

[16]:
```
     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
..                 …                 …                  …                 …
145                6.7               3.0                5.2               2.3
146                6.3               2.5                5.0               1.9
147                6.5               3.0                5.2               2.0
148                6.2               3.4                5.4               2.3
149                5.9               3.0                5.1               1.8

              target
0        Iris-setosa
1        Iris-setosa
2        Iris-setosa
3        Iris-setosa
4        Iris-setosa
..               …
145  Iris-virginica
146  Iris-virginica
147  Iris-virginica
148  Iris-virginica
149  Iris-virginica

[150 rows x 5 columns]
```

[17]:
```
hari.isnull().sum()
```

[17]:
```
sepal length (cm)    0
sepal width (cm)     0
petal length (cm)    0
petal width (cm)     0
target               0
dtype: int64
```

[18]:
```
hari.describe()
```

[18]:
```
       sepal length (cm)  sepal width (cm)  petal length (cm)  \
count         150.000000        150.000000         150.000000
mean            5.843333          3.054000           3.758667
std             0.828066          0.433594           1.764420
min             4.300000          2.000000           1.000000
25%             5.100000          2.800000           1.600000
50%             5.800000          3.000000           4.350000
75%             6.400000          3.300000           5.100000
max             7.900000          4.400000           6.900000

       petal width (cm)
count        150.000000
mean           1.198667
std            0.763161
min            0.100000
25%            0.300000
50%            1.300000
75%            1.800000
max            2.500000
```

[33]:
```
df.columns = df.columns.str.strip().str.replace(" ", "_").str.lower()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[33], line 1
----> 1 df.columns = df.columns.str.strip().str.replace(" ", "_").str.lower()

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

[19]:
```
hari.describe(include="all")
```

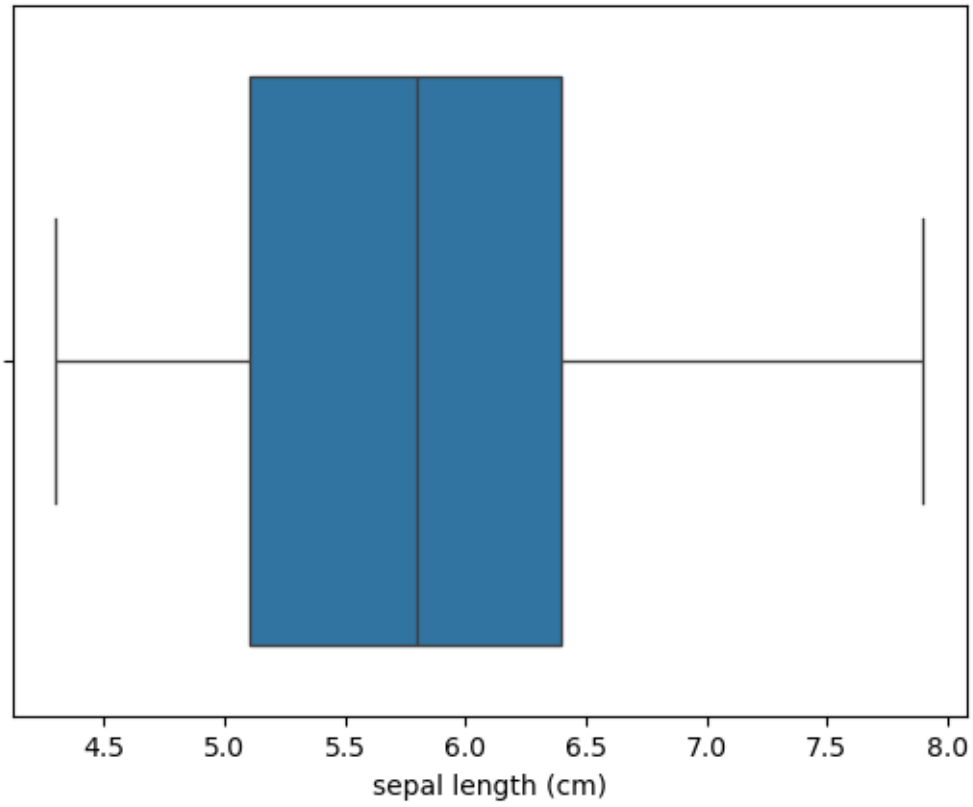
[19]:
```
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
count          150.000000        150.000000         150.000000
unique                NaN               NaN                NaN
top                   NaN               NaN                NaN
freq                  NaN               NaN                NaN
mean             5.843333          3.054000           3.758667
```

```
std              0.828066          0.433594          1.764420
min              4.300000          2.000000          1.000000
25%              5.100000          2.800000          1.600000
50%              5.800000          3.000000          4.350000
75%              6.400000          3.300000          5.100000
max              7.900000          4.400000          6.900000

        petal width (cm)       target
count         150.000000          150
unique               NaN            3
top                  NaN  Iris-setosa
freq                 NaN           50
mean            1.198667          NaN
std             0.763161          NaN
min             0.100000          NaN
25%             0.300000          NaN
50%             1.300000          NaN
75%             1.800000          NaN
max             2.500000          NaN
```

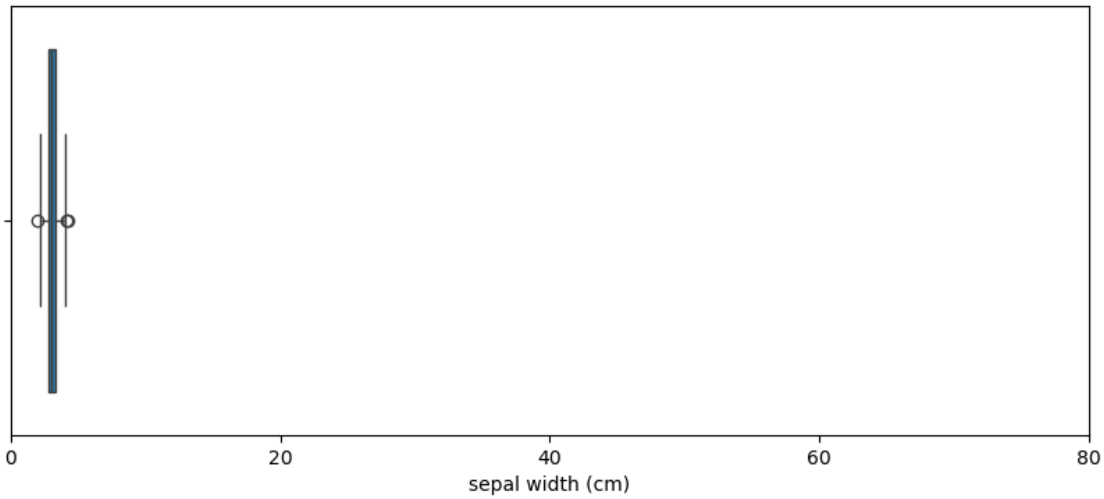
[38]:
```
sns.boxplot(x="sepal length (cm)",data=hari)
```

[38]:
```
<Axes: xlabel='sepal length (cm)'>
```



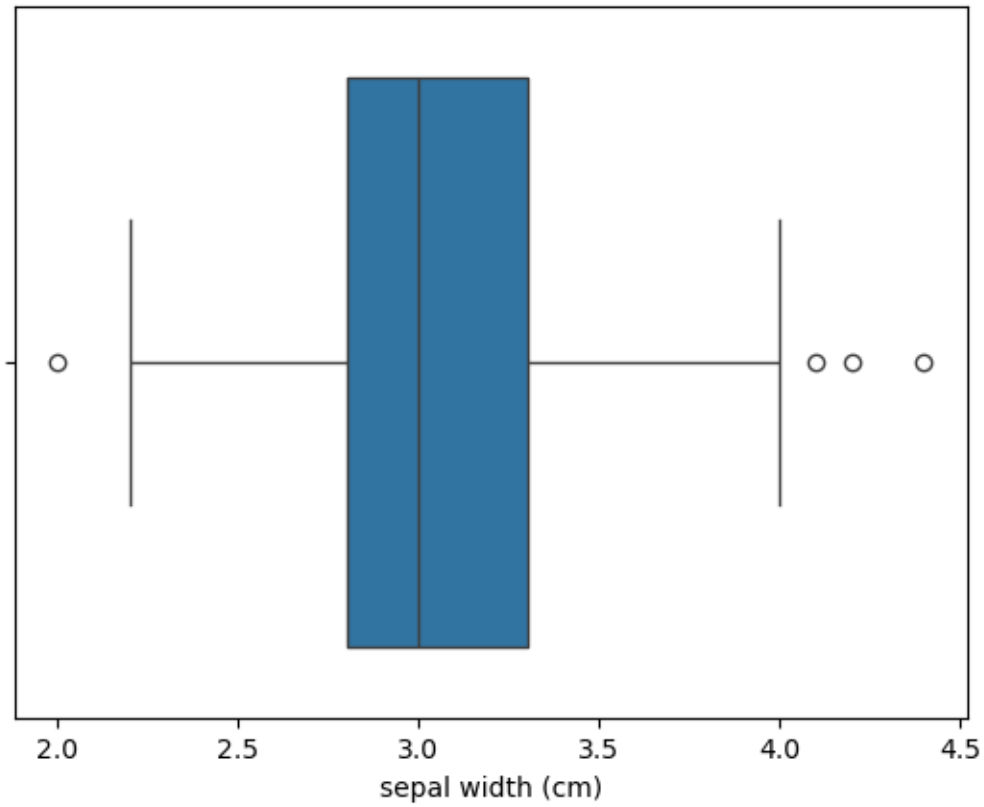
[22]:
```
def trim(col:pd.core.series.Series):
     lowerlimit=col.mean()-3*col.std()
     upperlimit=col.mean()+3*col.std()
     new_col = col[(col<upperlimit)&(col>lowerlimit)]
     return new_col
```

[39]:
```
plt.figure(figsize=(10,4))
sns.boxplot(x=trim(hari["sepal width (cm)"]),data=hari)
plt.xticks(np.arange(0,100,20))
plt.show()
```



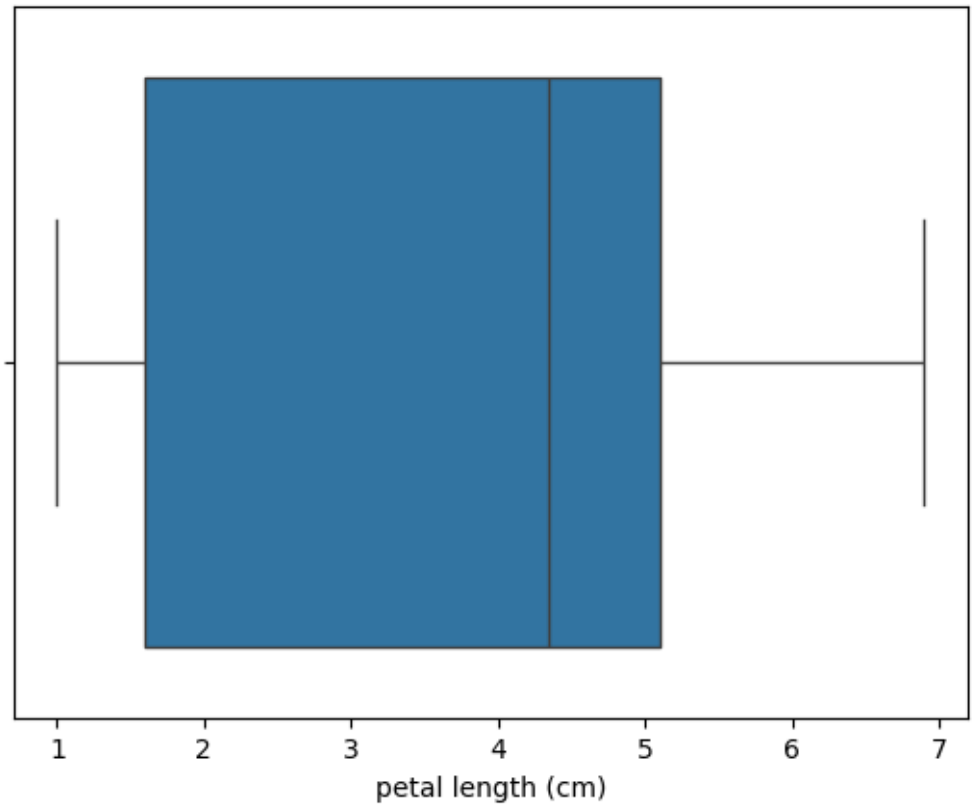
[40]:
```
sns.boxplot(x="sepal width (cm)",data=hari)
```

[40]:
```
<Axes: xlabel='sepal width (cm)'>
```



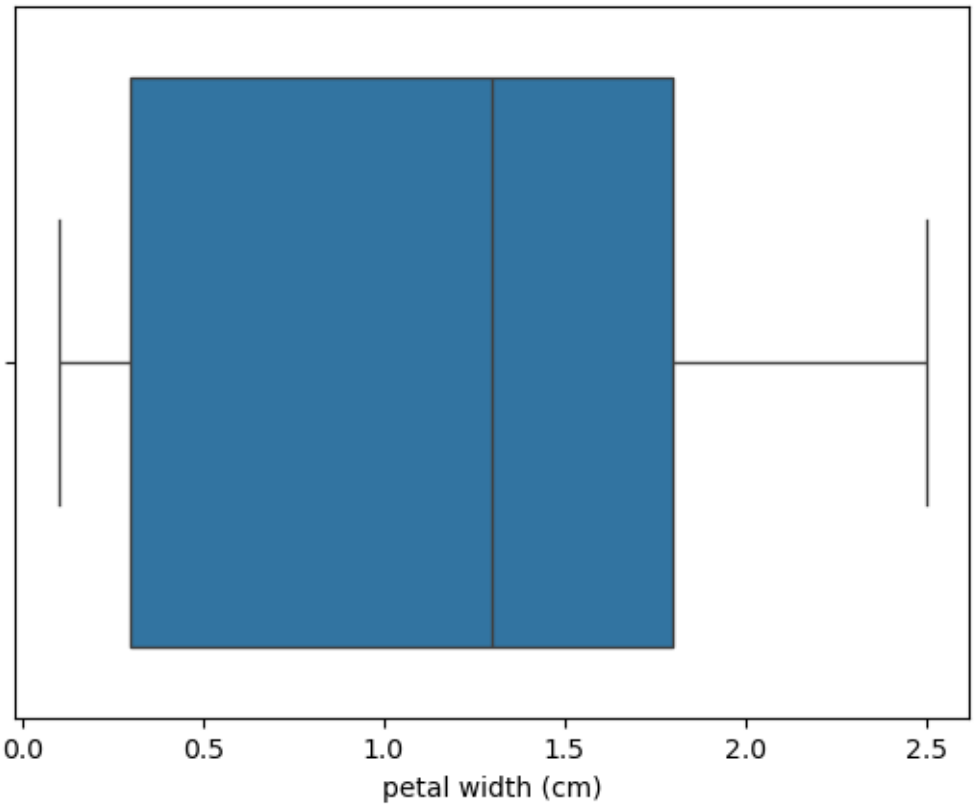
[37]:
```
sns.boxplot(x="petal length (cm)",data=hari)
```

[37]:
```
<Axes: xlabel='petal length (cm)'>
```



[41]:
```
sns.boxplot(x="petal width (cm)",data=hari)
```

[41]:
```
<Axes: xlabel='petal width (cm)'>
```



[43]:
```
hari.target.unique()
```

[43]:
```
array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

[46]:
```
# converting categorical veriable to numeric veriable
from sklearn.preprocessing import LabelEncoder
```

[47]:
```
le=LabelEncoder()
```

[50]:
```
le.fit_transform(hari["target"])
```

[50]:
```
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, 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])
```

[51]:
```
hari["target"]=le.fit_transform(hari["target"])
```

[52]:
```
hari
```

[52]:
```
     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
..                 …                 …                  …                 …
145                6.7               3.0                5.2               2.3
146                6.3               2.5                5.0               1.9
147                6.5               3.0                5.2               2.0
148                6.2               3.4                5.4               2.3
149                5.9               3.0                5.1               1.8

     target
0         0
1         0
2         0
3         0
4         0
..      …
145       2
146       2
147       2
148       2
149       2

[150 rows x 5 columns]
```

[54]:
```
#group data
hari.target.unique()
```

[54]:
```
array([0, 1, 2])
```

[55]:
```
g= hari.groupby("target")
```

[56]:
```
for hari,target_hari in g:
    print(hari)
    print(".............................")
    print(target_hari)
```

```
0
…
    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
5	5.4	3.9	1.7	0.4
6	4.6	3.4	1.4	0.3
7	5.0	3.4	1.5	0.2
8	4.4	2.9	1.4	0.2
9	4.9	3.1	1.5	0.1
10	5.4	3.7	1.5	0.2
11	4.8	3.4	1.6	0.2
12	4.8	3.0	1.4	0.1
13	4.3	3.0	1.1	0.1
14	5.8	4.0	1.2	0.2
15	5.7	4.4	1.5	0.4
16	5.4	3.9	1.3	0.4
17	5.1	3.5	1.4	0.3
18	5.7	3.8	1.7	0.3
19	5.1	3.8	1.5	0.3
20	5.4	3.4	1.7	0.2
21	5.1	3.7	1.5	0.4
22	4.6	3.6	1.0	0.2
23	5.1	3.3	1.7	0.5
24	4.8	3.4	1.9	0.2
25	5.0	3.0	1.6	0.2
26	5.0	3.4	1.6	0.4
27	5.2	3.5	1.5	0.2
28	5.2	3.4	1.4	0.2
29	4.7	3.2	1.6	0.2
30	4.8	3.1	1.6	0.2
31	5.4	3.4	1.5	0.4
32	5.2	4.1	1.5	0.1
33	5.5	4.2	1.4	0.2
34	4.9	3.1	1.5	0.1
35	5.0	3.2	1.2	0.2
36	5.5	3.5	1.3	0.2
37	4.9	3.1	1.5	0.1
38	4.4	3.0	1.3	0.2
39	5.1	3.4	1.5	0.2
40	5.0	3.5	1.3	0.3
41	4.5	2.3	1.3	0.3
42	4.4	3.2	1.3	0.2
43	5.0	3.5	1.6	0.6
44	5.1	3.8	1.9	0.4
45	4.8	3.0	1.4	0.3
46	5.1	3.8	1.6	0.2
47	4.6	3.2	1.4	0.2
48	5.3	3.7	1.5	0.2
49	5.0	3.3	1.4	0.2

	target
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0
9	0
10	0
11	0
12	0
13	0
14	0
15	0
16	0
17	0
18	0
19	0
20	0
21	0
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0
34	0
35	0
36	0
37	0
38	0
39	0
40	0
41	0
42	0
43	0
44	0
45	0
46	0

```
47       0
48       0
49       0
1
…
    sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
50                7.0               3.2                4.7               1.4
51                6.4               3.2                4.5               1.5
52                6.9               3.1                4.9               1.5
53                5.5               2.3                4.0               1.3
54                6.5               2.8                4.6               1.5
55                5.7               2.8                4.5               1.3
56                6.3               3.3                4.7               1.6
57                4.9               2.4                3.3               1.0
58                6.6               2.9                4.6               1.3
59                5.2               2.7                3.9               1.4
60                5.0               2.0                3.5               1.0
61                5.9               3.0                4.2               1.5
62                6.0               2.2                4.0               1.0
63                6.1               2.9                4.7               1.4
64                5.6               2.9                3.6               1.3
65                6.7               3.1                4.4               1.4
66                5.6               3.0                4.5               1.5
67                5.8               2.7                4.1               1.0
68                6.2               2.2                4.5               1.5
69                5.6               2.5                3.9               1.1
70                5.9               3.2                4.8               1.8
71                6.1               2.8                4.0               1.3
72                6.3               2.5                4.9               1.5
73                6.1               2.8                4.7               1.2
74                6.4               2.9                4.3               1.3
75                6.6               3.0                4.4               1.4
76                6.8               2.8                4.8               1.4
77                6.7               3.0                5.0               1.7
78                6.0               2.9                4.5               1.5
79                5.7               2.6                3.5               1.0
80                5.5               2.4                3.8               1.1
81                5.5               2.4                3.7               1.0
82                5.8               2.7                3.9               1.2
83                6.0               2.7                5.1               1.6
84                5.4               3.0                4.5               1.5
85                6.0               3.4                4.5               1.6
86                6.7               3.1                4.7               1.5
87                6.3               2.3                4.4               1.3
88                5.6               3.0                4.1               1.3
89                5.5               2.5                4.0               1.3
90                5.5               2.6                4.4               1.2
91                6.1               3.0                4.6               1.4
```

92	5.8	2.6	4.0	1.2
93	5.0	2.3	3.3	1.0
94	5.6	2.7	4.2	1.3
95	5.7	3.0	4.2	1.2
96	5.7	2.9	4.2	1.3
97	6.2	2.9	4.3	1.3
98	5.1	2.5	3.0	1.1
99	5.7	2.8	4.1	1.3

	target
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1

```
88       1
89       1
90       1
91       1
92       1
93       1
94       1
95       1
96       1
97       1
98       1
99       1
2
…
     sepal length (cm)  sepal width (cm)  petal length (cm)  petal width (cm)  \
100                6.3               3.3                6.0               2.5
101                5.8               2.7                5.1               1.9
102                7.1               3.0                5.9               2.1
103                6.3               2.9                5.6               1.8
104                6.5               3.0                5.8               2.2
105                7.6               3.0                6.6               2.1
106                4.9               2.5                4.5               1.7
107                7.3               2.9                6.3               1.8
108                6.7               2.5                5.8               1.8
109                7.2               3.6                6.1               2.5
110                6.5               3.2                5.1               2.0
111                6.4               2.7                5.3               1.9
112                6.8               3.0                5.5               2.1
113                5.7               2.5                5.0               2.0
114                5.8               2.8                5.1               2.4
115                6.4               3.2                5.3               2.3
116                6.5               3.0                5.5               1.8
117                7.7               3.8                6.7               2.2
118                7.7               2.6                6.9               2.3
119                6.0               2.2                5.0               1.5
120                6.9               3.2                5.7               2.3
121                5.6               2.8                4.9               2.0
122                7.7               2.8                6.7               2.0
123                6.3               2.7                4.9               1.8
124                6.7               3.3                5.7               2.1
125                7.2               3.2                6.0               1.8
126                6.2               2.8                4.8               1.8
127                6.1               3.0                4.9               1.8
128                6.4               2.8                5.6               2.1
129                7.2               3.0                5.8               1.6
130                7.4               2.8                6.1               1.9
131                7.9               3.8                6.4               2.0
132                6.4               2.8                5.6               2.2
```

133	6.3	2.8	5.1	1.5
134	6.1	2.6	5.6	1.4
135	7.7	3.0	6.1	2.3
136	6.3	3.4	5.6	2.4
137	6.4	3.1	5.5	1.8
138	6.0	3.0	4.8	1.8
139	6.9	3.1	5.4	2.1
140	6.7	3.1	5.6	2.4
141	6.9	3.1	5.1	2.3
142	5.8	2.7	5.1	1.9
143	6.8	3.2	5.9	2.3
144	6.7	3.3	5.7	2.5
145	6.7	3.0	5.2	2.3
146	6.3	2.5	5.0	1.9
147	6.5	3.0	5.2	2.0
148	6.2	3.4	5.4	2.3
149	5.9	3.0	5.1	1.8

	target
100	2
101	2
102	2
103	2
104	2
105	2
106	2
107	2
108	2
109	2
110	2
111	2
112	2
113	2
114	2
115	2
116	2
117	2
118	2
119	2
120	2
121	2
122	2
123	2
124	2
125	2
126	2
127	2
128	2

```
129       2
130       2
131       2
132       2
133       2
134       2
135       2
136       2
137       2
138       2
139       2
140       2
141       2
142       2
143       2
144       2
145       2
146       2
147       2
148       2
149       2
```

[57]:
```
#stasics
g.agg({'sepal length (cm)':'mean'})
```

[57]:
```
        sepal length (cm)
target
0                   5.006
1                   5.936
2                   6.588
```

[59]:
```
g.agg ({"sepal length (cm)":"mean","sepal width (cm)":"mean","petal length␣
↪(cm)":"mean", "petal width (cm)":"mean"})
```

[59]:
```
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
target
0                   5.006             3.418              1.464
1                   5.936             2.770              4.260
2                   6.588             2.974              5.552

        petal width (cm)
target
0                  0.244
1                  1.326
2                  2.026
```

[60]:
```
print ("Mean of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":"mean","sepal width (cm)":"mean","petal length␣
↪(cm)":"mean", "petal width (cm)":"mean"})
```

```
Mean of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[60]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.006	3.418	1.464	0.244
1	5.936	2.770	4.260	1.326
2	6.588	2.974	5.552	2.026

[61]:
```
print ("Median of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":"median","sepal width (cm)":"median","petal length␣
↪(cm)":"median", "petal width (cm)":"median"})
```

```
Median of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[61]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.0	3.4	1.50	0.2
1	5.9	2.8	4.35	1.3
2	6.5	3.0	5.55	2.0

[64]:
```
print ("std of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":"std","sepal width (cm)":"std","petal length (cm)":
 ↪"std", "petal width (cm)":"std"})
```

```
std of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[64]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	0.352490	0.381024	0.173511	0.107210
1	0.516171	0.313798	0.469911	0.197753
2	0.635880	0.322497	0.551895	0.274650

[71]:
```
print ("25 percentile of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":lambda x:x.quantile(0.25),"sepal width (cm)":lambda␣
 ↪x:x.quantile(0.25),"petal length (cm)":lambda x:x.quantile(0.25),"petal␣
 ↪width (cm)":lambda x:x.quantile(0.25)})
```

```
25 percentile of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[71]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	4.800	3.125	1.4	0.2
1	5.600	2.525	4.0	1.2
2	6.225	2.800	5.1	1.8

[72]:
```
print ("50 percentile of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":lambda x:x.quantile(0.50),"sepal width (cm)":
↪lambda x:x.quantile(0.50),"petal length (cm)":lambda x:x.quantile(0.
↪50),"petal width (cm)":lambda x:x.quantile(0.50)})
```

```
50 percentile of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[72]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.0	3.4	1.50	0.2
1	5.9	2.8	4.35	1.3
2	6.5	3.0	5.55	2.0

[74]:
```
print ("75 percentile of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":lambda x:x.quantile(0.75),"sepal width (cm)":lambda␣
↪x:x.quantile(0.75),"petal length (cm)":lambda x:x.quantile(0.75),"petal␣
↪width (cm)":lambda x:x.quantile(0.75)})
```

```
75 percentile of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[74]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.2	3.675	1.575	0.3
1	6.3	3.000	4.600	1.5
2	6.9	3.175	5.875	

```
2                   2.3
```

[75]:
```
print ("min of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":"min","sepal width (cm)":"min","petal length (cm)":
↪"min", "petal width (cm)":"min"})
```

```
min of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[75]:
```
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
target
0                     4.3               2.3                1.0
1                     4.9               2.0                3.0
2                     4.9               2.2                4.5

        petal width (cm)
target
0                    0.1
1                    1.0
2                    1.4
```

[76]:
```
print ("max of Eeach target")
print("0=Iris Setosa")
print("1=Iris Versicolor")
print("2=Iris Verginia")
g.agg ({"sepal length (cm)":"max","sepal width (cm)":"max","petal length (cm)":
↪"max", "petal width (cm)":"max"})
```

```
max of Eeach target
0=Iris Setosa
1=Iris Versicolor
2=Iris Verginia
```

[76]:
```
        sepal length (cm)  sepal width (cm)  petal length (cm)  \
target
0                     5.8               4.4                1.9
1                     7.0               3.4                5.1
2                     7.9               3.8                6.9

        petal width (cm)
target
0                    0.6
1                    1.8
```

```
2                    2.5
```

[83]:
```
hari1=pd.read_csv('headbrain.csv')
```

[84]:
```
hari1
```

[84]:
```
     Gender  Age Range  Head Size(cm^3)  Brain Weight(grams)
0         1          1             4512                 1530
1         1          1             3738                 1297
2         1          1             4261                 1335
3         1          1             3777                 1282
4         1          1             4177                 1590
..      …          …                …                    …
232       2          2             3214                 1110
233       2          2             3394                 1215
234       2          2             3233                 1104
235       2          2             3352                 1170
236       2          2             3391                 1120

[237 rows x 4 columns]
```

[85]:
```
hari1.skew(axis = 0, skipna = True)
```

[85]:
```
Gender                 0.265554
Age Range             -0.144748
Head Size(cm^3)        0.238807
Brain Weight(grams)    0.265478
dtype: float64
```

[86]:
```
hari1.skew(axis = 1, skipna = True)
```

[86]:
```
0      1.396270
1      1.370942
2      1.477415
3      1.395312
4      1.259623
         …
232    1.376919
233    1.335684
234    1.389454
235    1.364923
236    1.425139
Length: 237, dtype: float64
```

[87]:
```
hari1
```

[87]:

	Gender	Age Range	Head Size(cm^3)	Brain Weight(grams)
0	1	1	4512	1530
1	1	1	3738	1297
2	1	1	4261	1335
3	1	1	3777	1282
4	1	1	4177	1590
..	…	…	…	…
232	2	2	3214	1110
233	2	2	3394	1215
234	2	2	3233	1104
235	2	2	3352	1170
236	2	2	3391	1120

[237 rows x 4 columns]

[ ]: