```
# Plot the correlation plot on dataset and visualize giving an
# overview of relationships among data on iris data.

import pandas as pd
iris=pd.read_csv("/content/drive/MyDrive/KRAI/iris.csv")
print(iris)
print(iris.head())
print(iris.tail())
print(iris.dtypes)
```

```
     sepal.length  sepal.width  petal.length  petal.width    variety
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  Virginica
146           6.3          2.5           5.0          1.9  Virginica
147           6.5          3.0           5.2          2.0  Virginica
148           6.2          3.4           5.4          2.3  Virginica
149           5.9          3.0           5.1          1.8  Virginica

[150 rows x 5 columns]
   sepal.length  sepal.width  petal.length  petal.width variety
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
     sepal.length  sepal.width  petal.length  petal.width    variety
145           6.7          3.0           5.2          2.3  Virginica
146           6.3          2.5           5.0          1.9  Virginica
147           6.5          3.0           5.2          2.0  Virginica
148           6.2          3.4           5.4          2.3  Virginica
149           5.9          3.0           5.1          1.8  Virginica
sepal.length    float64
sepal.width     float64
petal.length    float64
petal.width     float64
variety          object
dtype: object
```

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
from sklearn import metrics
sns.set()
```

```
iris_data=pd.read_csv("/content/drive/MyDrive/KRAI/iris.csv")
print(iris_data)
```

```
     sepal.length  sepal.width  petal.length  petal.width    variety
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  Virginica
146           6.3          2.5           5.0          1.9  Virginica
147           6.5          3.0           5.2          2.0  Virginica
148           6.2          3.4           5.4          2.3  Virginica
149           5.9          3.0           5.1          1.8  Virginica

[150 rows x 5 columns]
```

```
iris_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal.length  150 non-null    float64
 1   sepal.width   150 non-null    float64
 2   petal.length  150 non-null    float64
 3   petal.width   150 non-null    float64
 4   variety       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```
iris_data.describe()
```

	sepal.length	sepal.width	petal.length	petal.width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
iris_data[iris_data.duplicated()]
```

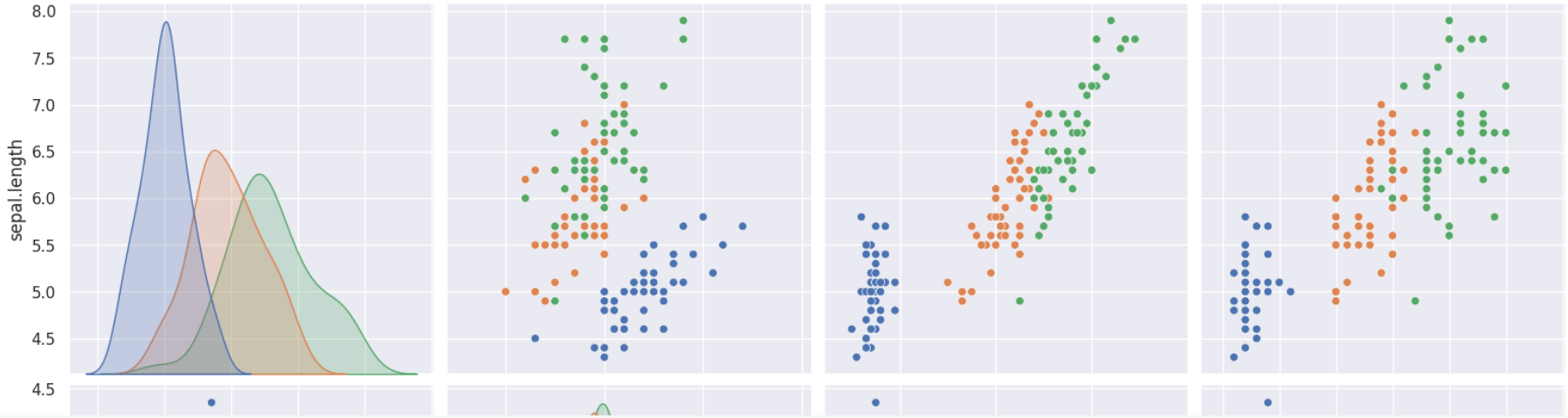
	sepal.length	sepal.width	petal.length	petal.width	variety
142	5.8	2.7	5.1	1.9	Virginica

```
iris_data['variety'].value_counts()
```

```
Setosa        50
Versicolor    50
Virginica     50
Name: variety, dtype: int64
```

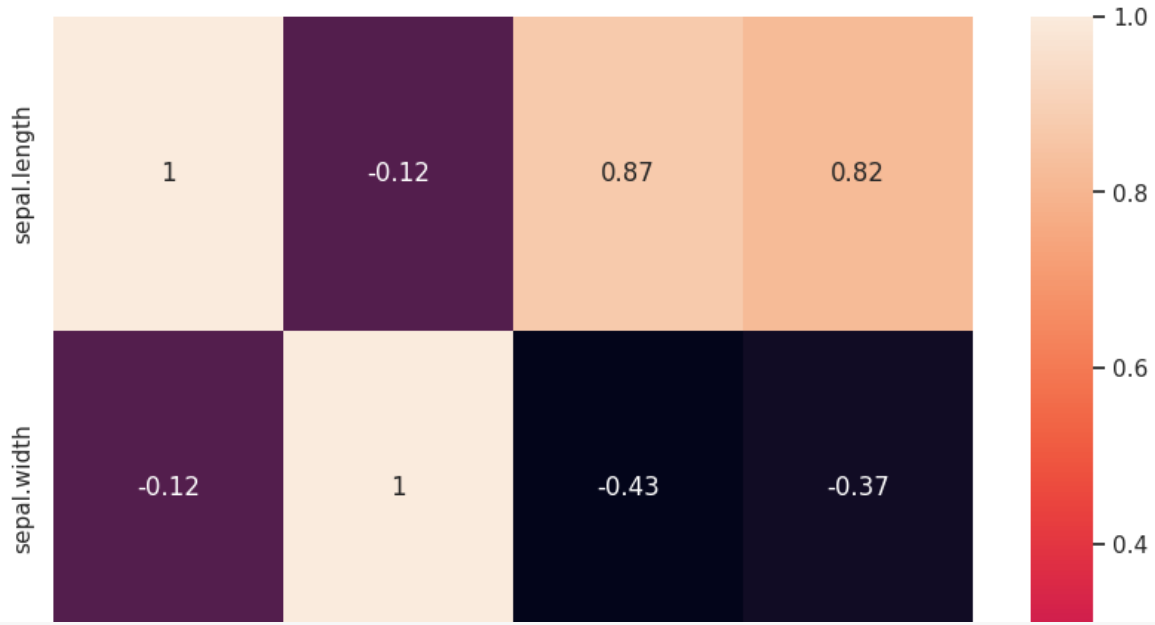
```
sns.pairplot(iris_data,hue='variety',height=4)
```

```
<seaborn.axisgrid.PairGrid at 0x7940b6d33370>
```



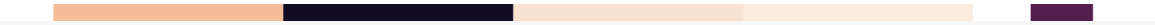

```
plt.figure(figsize=(10,11))
sns.heatmap(iris_data.corr(),annot=True)
plt.plot()
```

```
<ipython-input-14-0a05fdd33f33>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only
  sns.heatmap(iris_data.corr(),annot=True)
[]
```

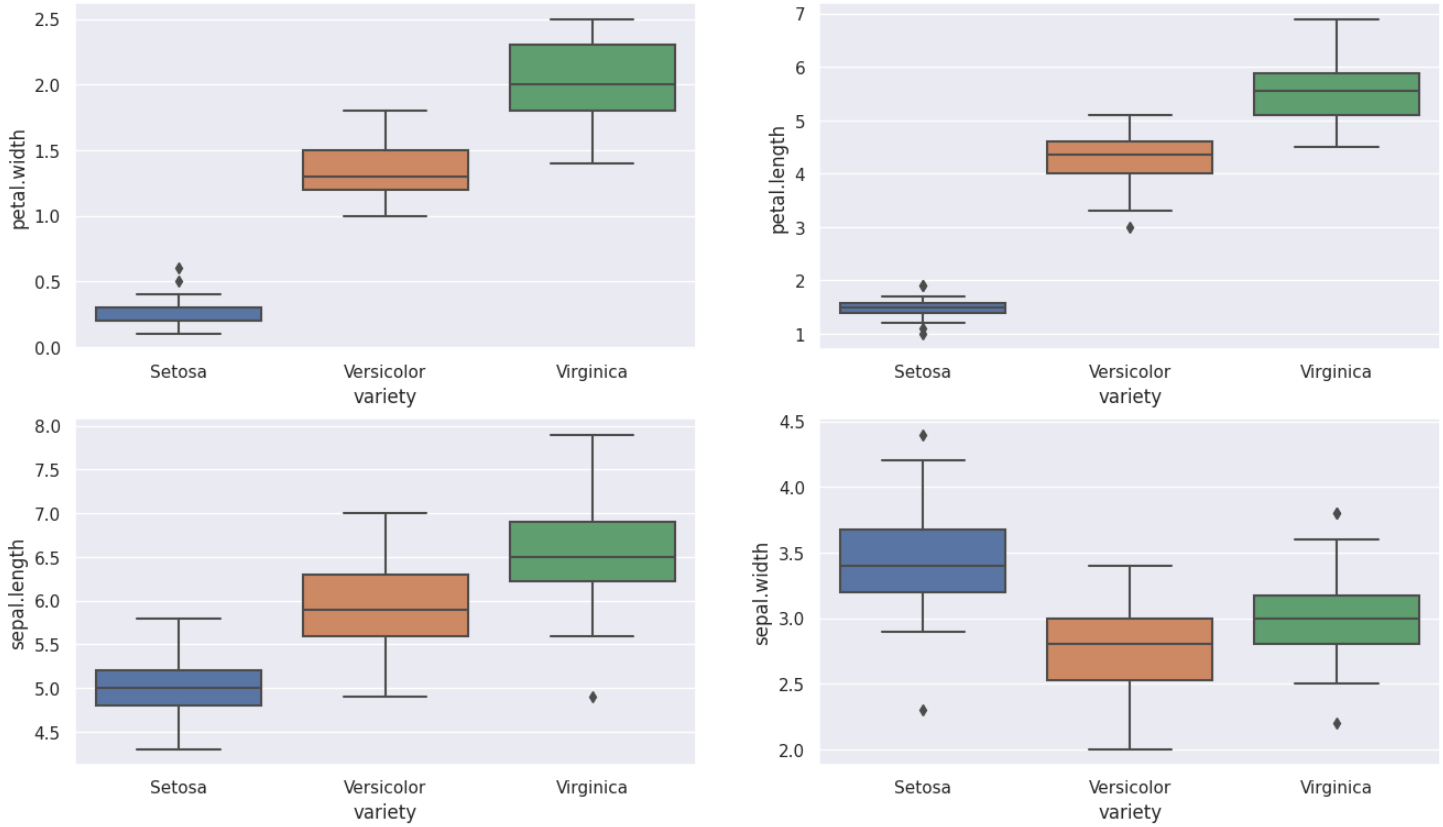



```
iris_data.groupby('variety').agg(['mean','median'])
```

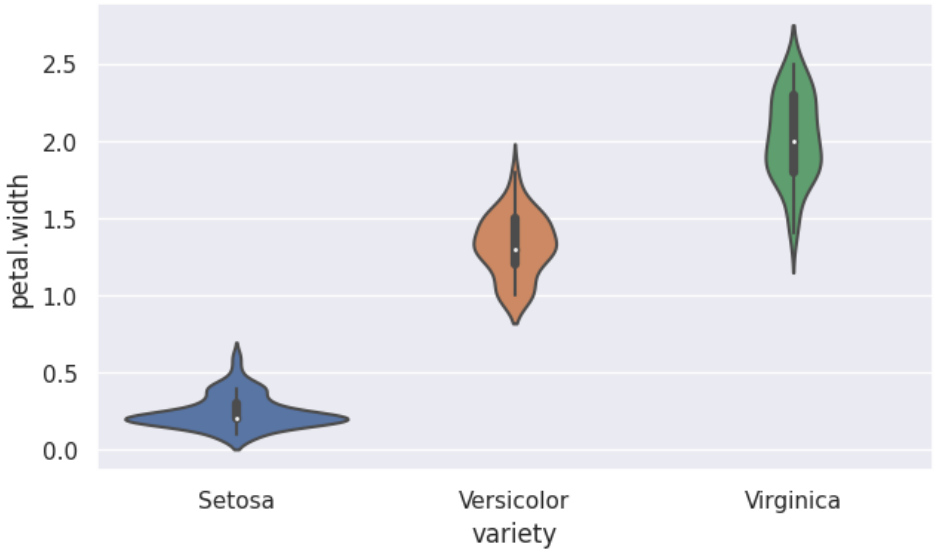
	sepal.length		sepal.width		petal.length		petal.width	
	mean	median	mean	median	mean	median	mean	median
variety								
Setosa	5.006	5.0	3.428	3.4	1.462	1.50	0.246	0.2
Versicolor	5.936	5.9	2.770	2.8	4.260	4.35	1.326	1.3
Virginica	6.588	6.5	2.974	3.0	5.552	5.55	2.026	2.0

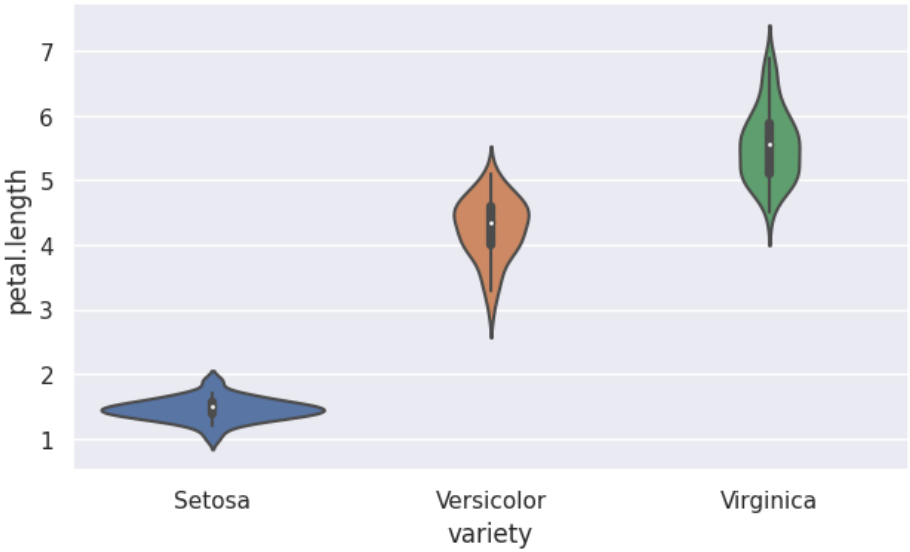


```
fig,axes=plt.subplots(2,2,figsize=(16,9))
sns.boxplot(y= 'petal.width' , x='variety' , data=iris_data , orient='v' , ax=axes[0,0])
sns.boxplot(y='petal.length',x='variety',data=iris_data,orient='v',ax=axes[0,1])
sns.boxplot(y='sepal.length',x='variety',data=iris_data,orient='v',ax=axes[1,0])
sns.boxplot(y='sepal.width',x='variety',data=iris_data,orient='v',ax=axes[1,1])
plt.show()
```




```
fig,axes=plt.subplots(2,2,figsize=(16,9))
sns.violinplot(y= 'petal.width', x='variety' , data=iris_data , orient='v' , ax=axes[0,0])
sns.violinplot(y='petal.length', x='variety' , data=iris_data , orient='v' , ax=axes[0,1])
sns.violinplot(y='sepal.length', x='variety' , data=iris_data , orient='v' , ax=axes[1,0])
sns.violinplot(y='sepal.width' , x='variety' , data=iris_data , orient='v' , ax=axes[1,1])
plt.show()
```






```
jupyter nbconvert --to pdf 2_prac_iris.ipynb
```

```
  File "<ipython-input-15-c903ee054443>", line 1
    jupyter nbconvert --to pdf 2_prac_iris.ipynb
                                  ^
SyntaxError: invalid decimal literal
```

SEARCH STACK OVERFLOW