

Iris Dataset :

jupyter TECHBEES (week 2) Last Checkpoint: Last Thursday at 5:39 PM (autosaved)



Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted

Python 3 (ipykernel)

Run Code

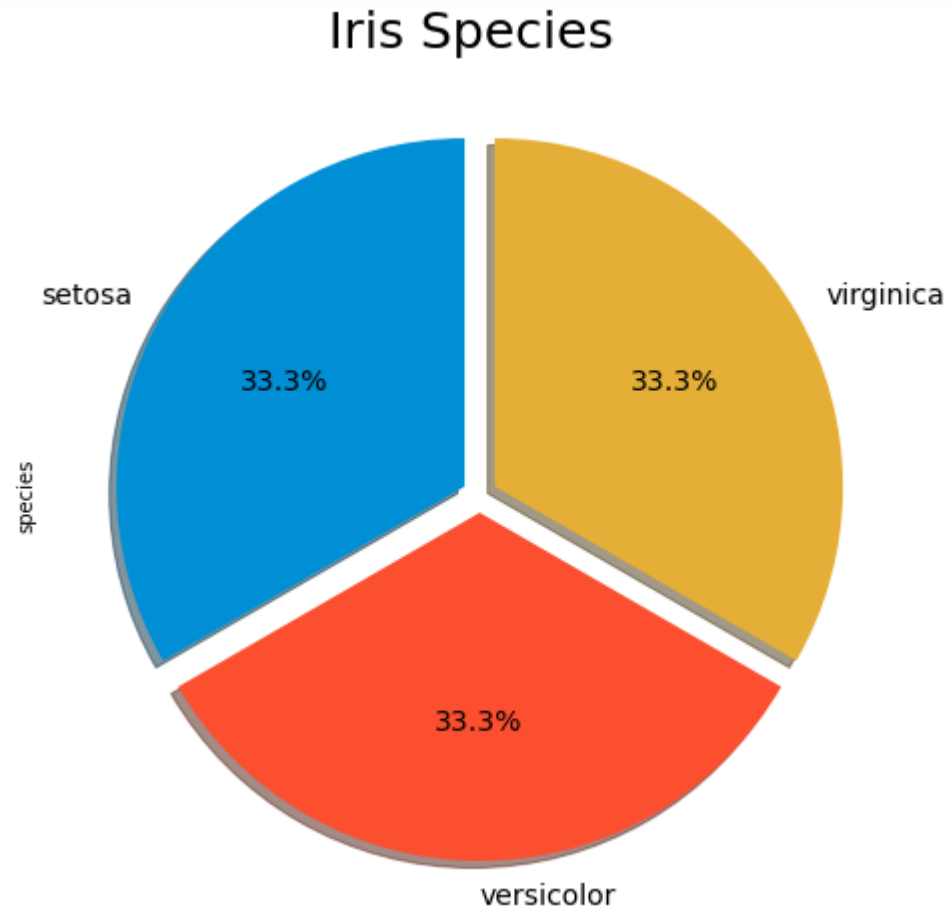
```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
data = sns.load_dataset('iris')
print(data)
```

	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	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]

1. Plot a pie chart of different species

```
: ax=plt.subplots(1,1,figsize=(10,8))
colors = ['#008fd5','#fc4f30','#e5ae37']
data['species'].value_counts().plot.pie(explode=[0.05,0.05,0.05],colors = colors,textprops={'fontsize': 14}
                                         ,autopct='%1.1f%%',startangle=90,shadow=True,figsize=(10,8))
plt.title("Iris Species ",fontsize=25)
plt.show()
```

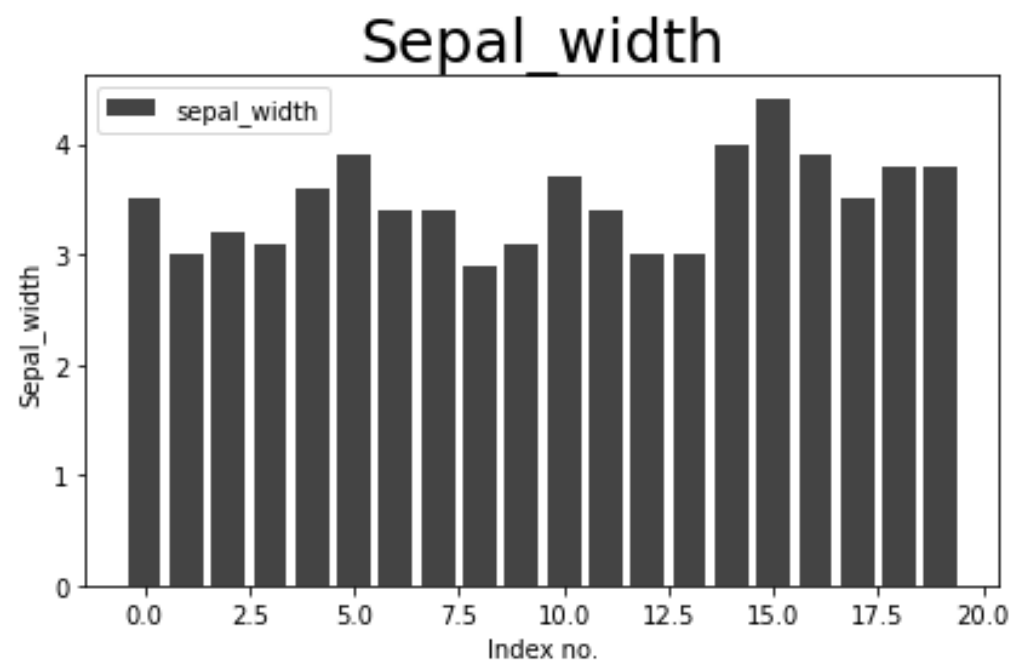


The insights of the 1st visualisation:

1. Setosa Species have 33.3% class distribution.
2. Virginica Species have 33.3% class distribution.
3. Versicolor Species have 33.3% class distribution.

2. Plot a bar graph of sepal_width of the first 20 records

```
df = data[['sepal_width']].head(20)
fig, ax = plt.subplots()
color = '#444444'
ax.bar(df.index, df['sepal_width'], label='sepal_width', color=color)
ax.set_title('Sepal_width', fontsize = 25)
ax.set_xlabel('Index no.')
ax.set_ylabel('Sepal_width')
plt.legend()
plt.tight_layout()
plt.show()
```

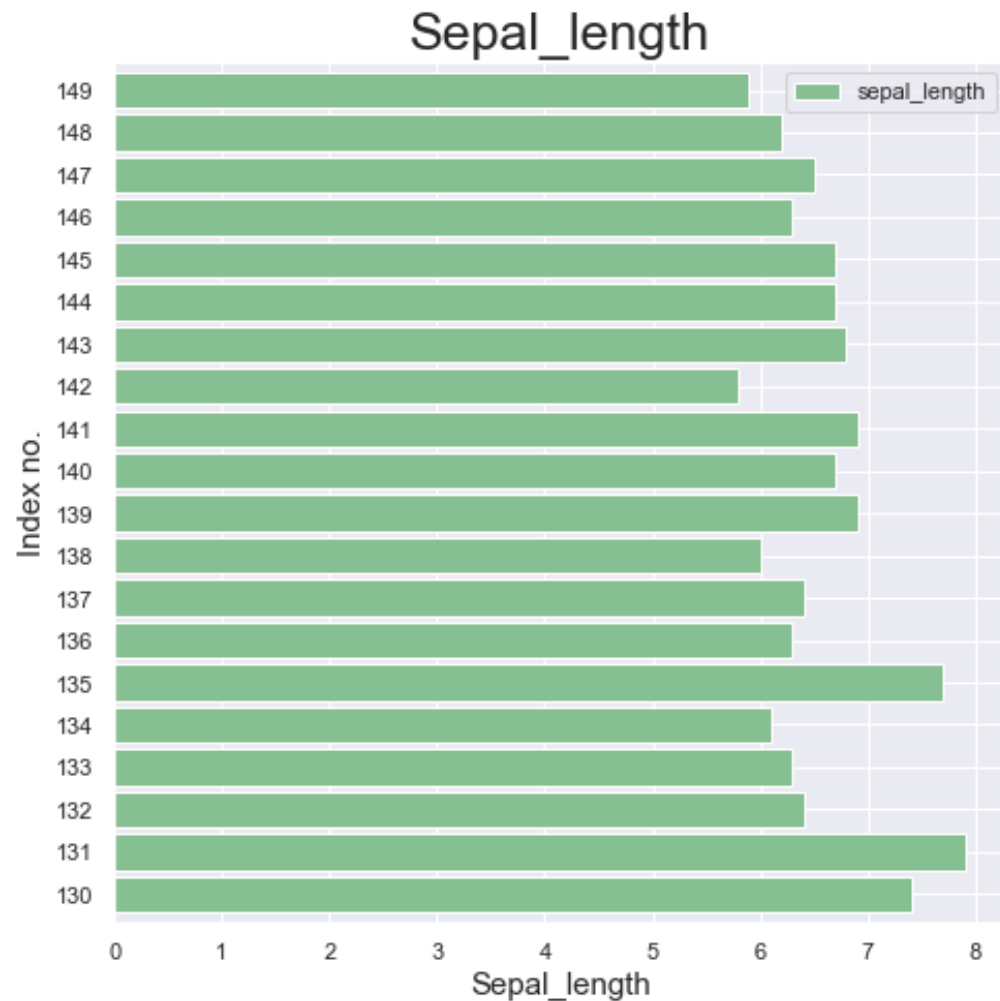


The insights of the 2nd visualisation:

1. Index number 15 has the highest Sepal Width which is more than 4 unit.
2. Index number 8 has the lowest Sepal Width which is less than 3 unit.
3. Most of the Sepal Width lies between 3 to 4 unit.

3. Plot a horizontal bar graph of sepal_length of the last 20 records.

```
df = data[['sepal_length']].tail(20)
ax = df.plot(kind='barh', figsize=(7, 7), color='#86bf91', width=0.85)
ax.set_title('Sepal_length',fontsize = 25)
ax.set_ylabel('Index no.',fontsize=15)
ax.set_xlabel('Sepal_length',fontsize=15)
plt.legend()
plt.tight_layout()
plt.show()
```

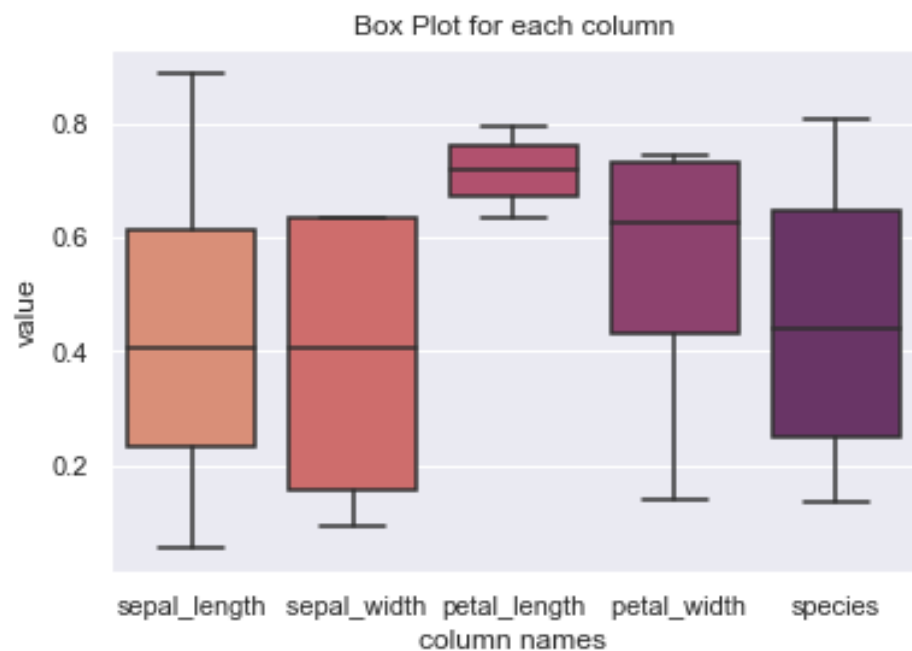


The insights of the 3rd visualisation:

1. Index number 131 has the highest Sepal Length which is more than 7.5 unit.
2. Index number 142 has the lowest Sepal Length which lies between 5.5 to 6 unit.
3. Most of the Sepal Length lies between 6 to 7 unit.

4. Plot a box plot for the each column using Seaborn Library, and find out the median for the column using the graph.

```
df = pd.DataFrame(data = np.random.random(size=(4,5)), columns =  
                  ['sepal_length', ' sepal_width', 'petal_length', ' petal_width', 'species'])  
  
sns.boxplot(x="variable", y="value", data=pd.melt(df),palette="flare").set(xlabel='column names',  
                                                                           ylabel='value',title='Box Plot for each column')  
  
plt.show()
```

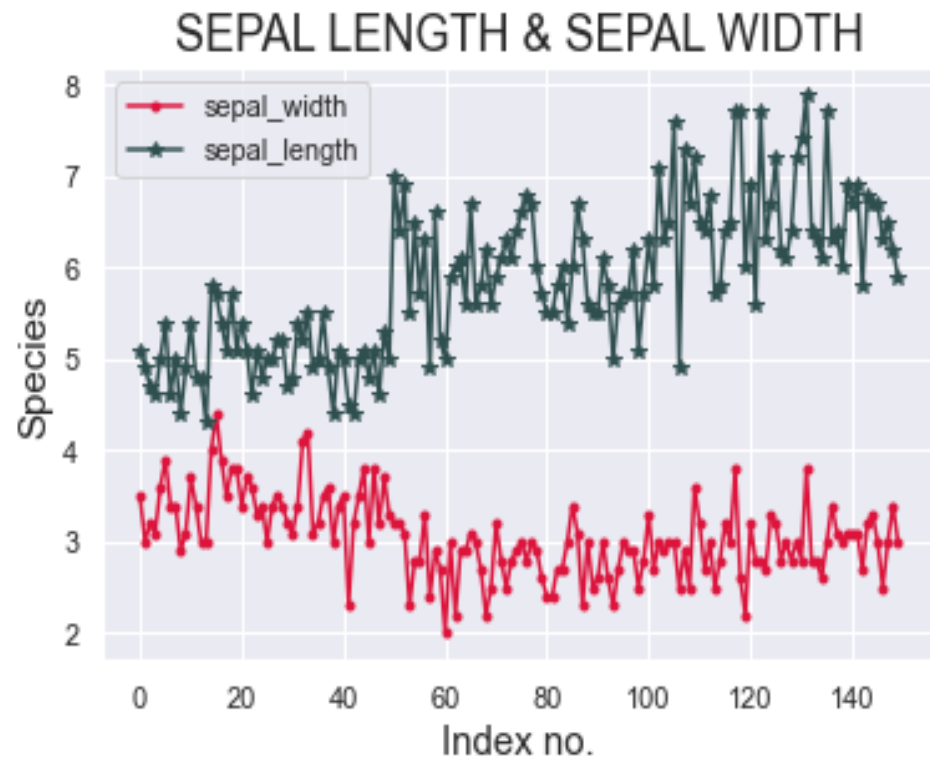


The insights of the 4th visualisation:

1. Petal Length has the maximum median.
2. Sepal Length and Sepal Width have the same median.
3. Sepal Length and Sepal Width have the lowest median.
4. Sepal Width has the highest value than the other columns.
5. Petal Length has the lowest value than the other columns.

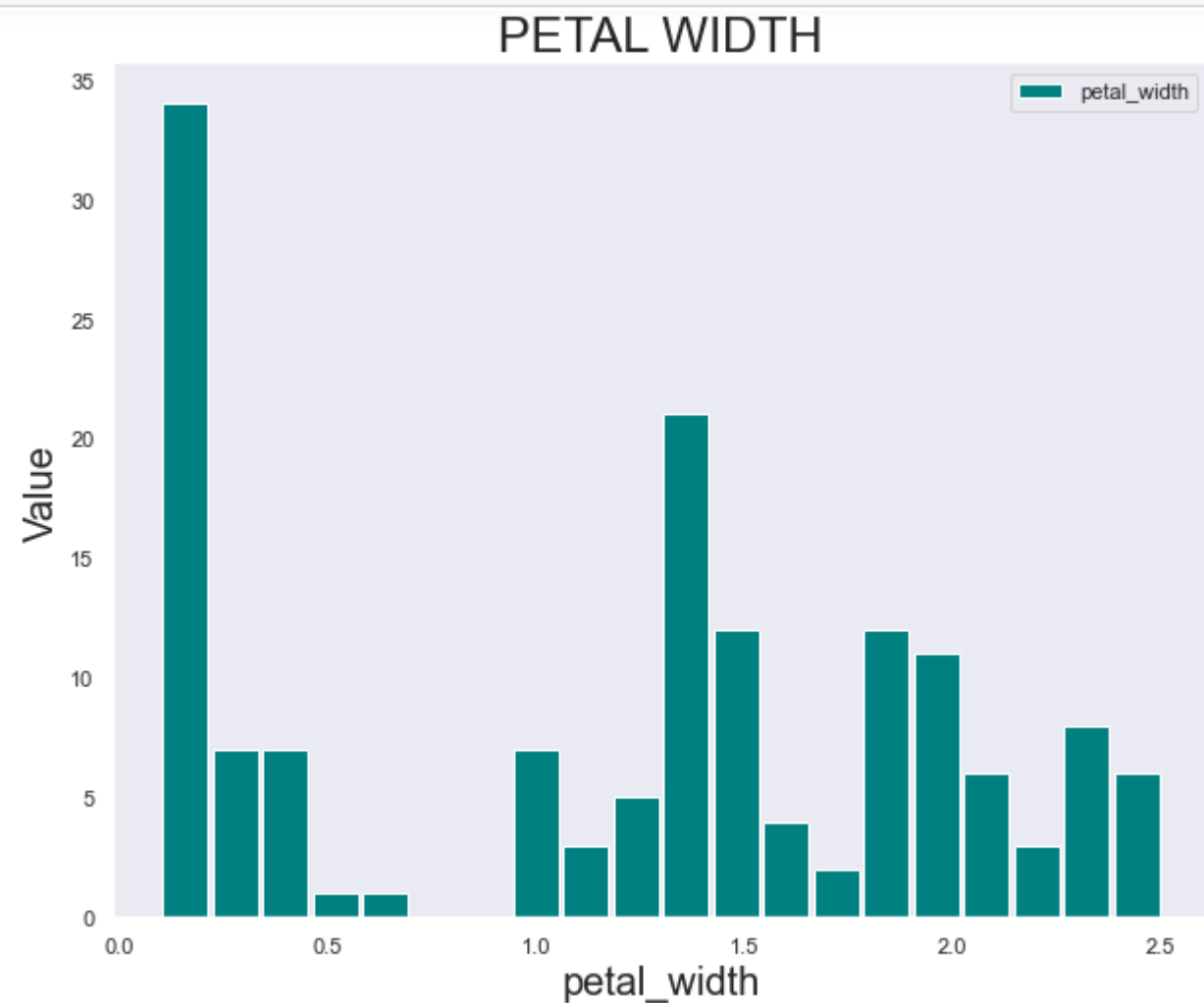
5. Plot a line chart containing both sepal_length and sepal_width

```
fig,ax = plt.subplots()
ax.plot(data.index,data['sepal_width'],label='sepal_width',color = 'crimson',marker='.')
ax.plot(data.index,data['sepal_length'],label='sepal_length',color = 'darkslategray',marker='*')
ax.set_xlabel('Index no.',fontsize=15)
ax.set_ylabel('Species',fontsize=15)
ax.set_title('SEPAL LENGTH & SEPAL WIDTH',fontsize=18)
plt.legend()
plt.show()
```



6. Plot a histogram for the column petal_width

```
ax = data.hist(column='petal_width', bins=20,grid=False ,figsize=(10,8), color='teal', zorder=2,  
               rwidth=0.9,label='petal_width')  
plt.xlabel('petal_width',fontsize=20)  
plt.ylabel('Value',fontsize=20)  
plt.title('PETAL WIDTH',fontsize=25)  
plt.legend()  
plt.show()
```

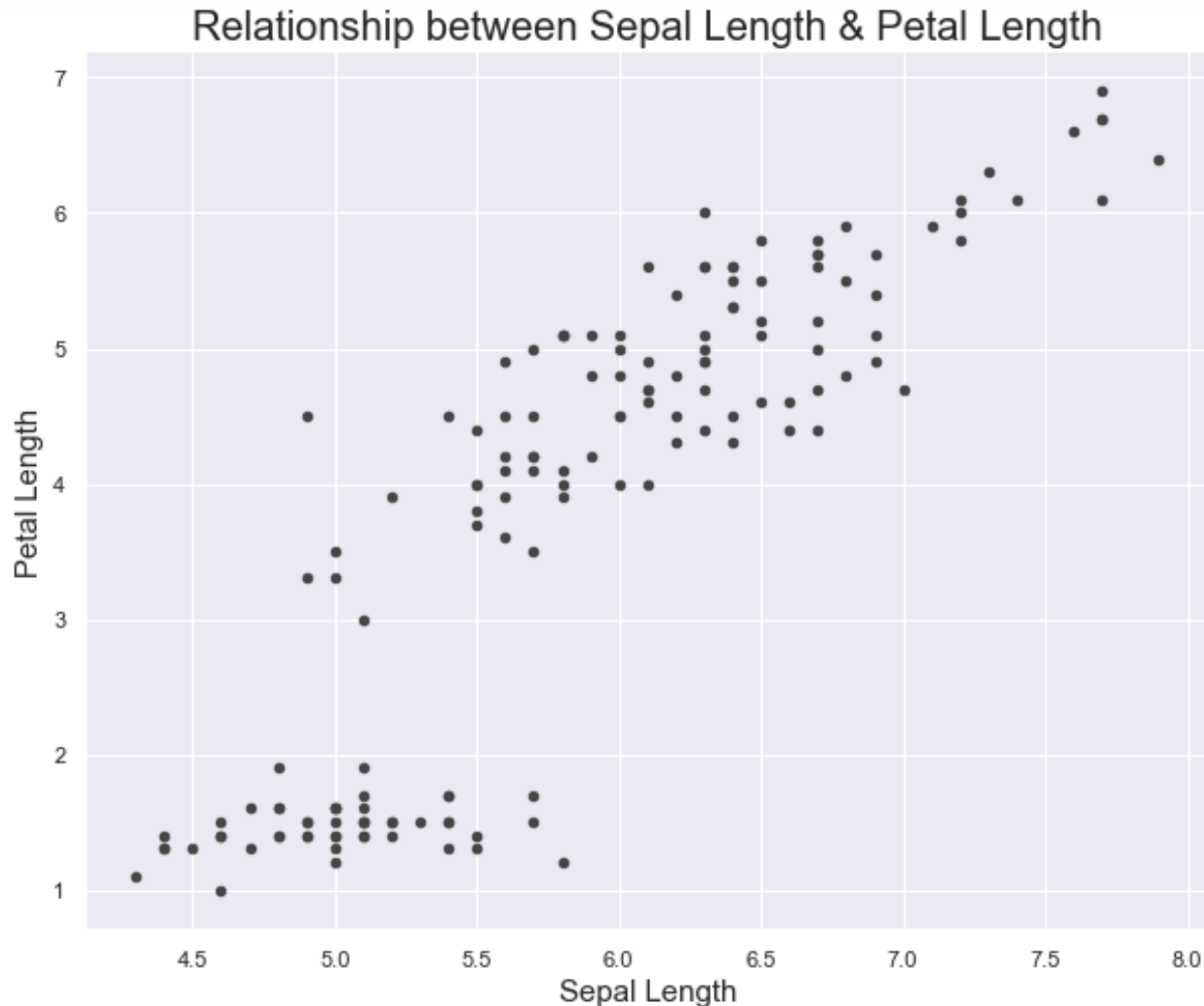


The insights of 6th visualisation:

1. There are largest number of 1.2 unit of petal width.
2. Petal width of 0.5 unit and 0.6 unit have the same value.
3. Most petal width value lies between 0 to 0.5 unit.

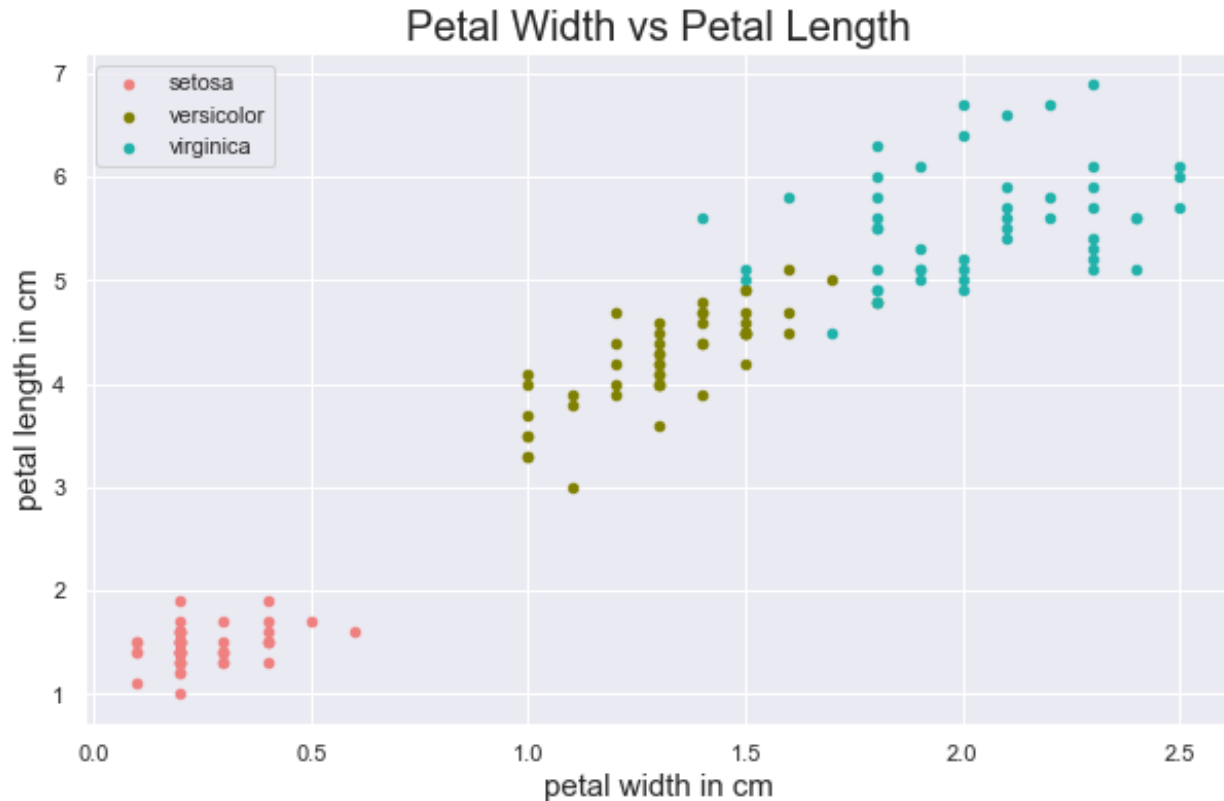
7. Plot a scatter plot to show relationship between sepal_length and petal_length

```
data.plot.scatter(x='sepal_length', y='petal_length',color='#444444',figsize=(10,8))  
plt.title("Relationship between Sepal Length & Petal Length",fontsize=20)  
plt.xlabel("Sepal Length",fontsize=15)  
plt.ylabel("Petal Length",fontsize=15)  
plt.show()
```



8. Plot a Seaborn Scatterplot to compare various species based on petal length and width.

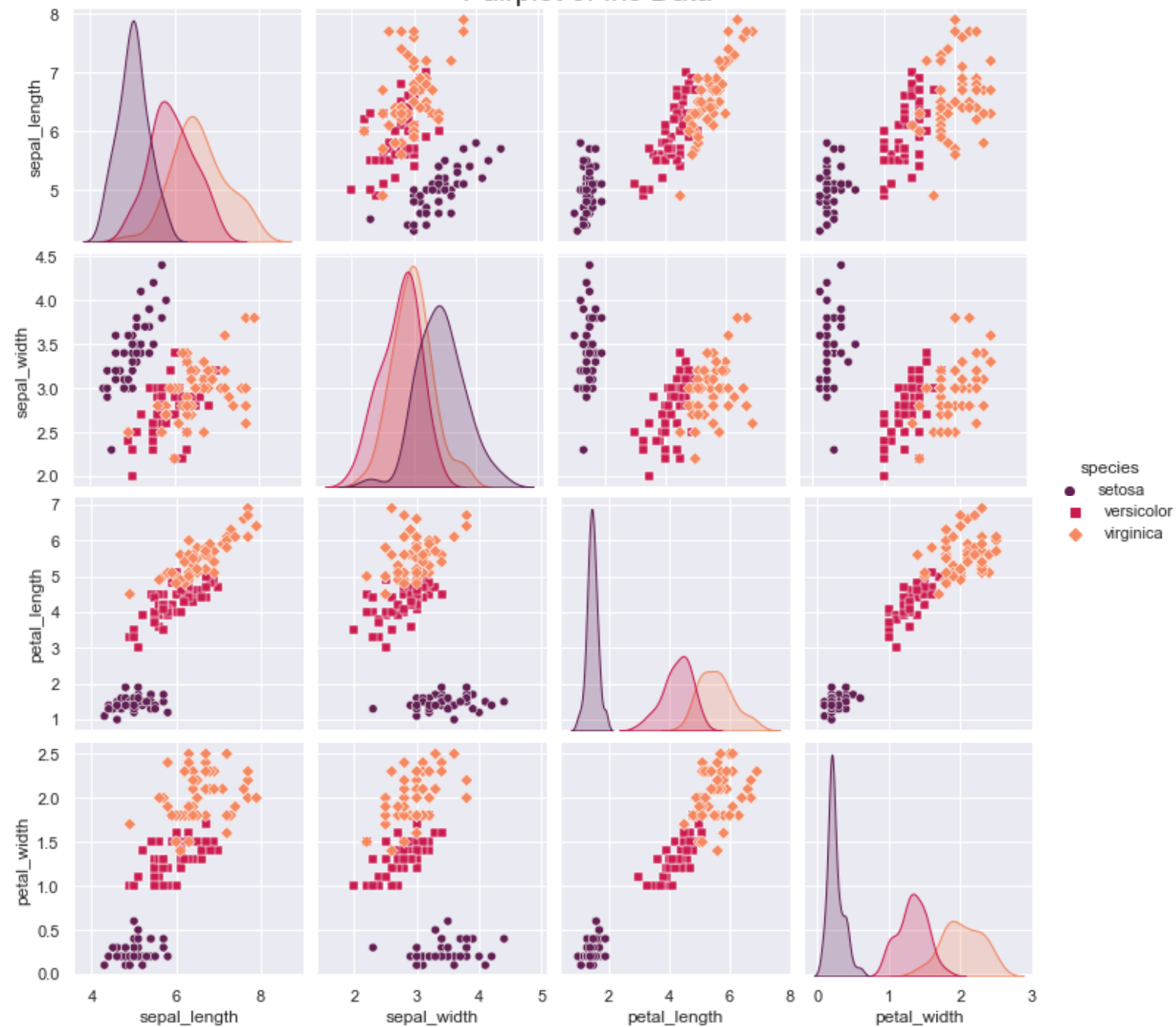
```
sns.set(style="darkgrid")
sc=data[data.species=='setosa'].plot(kind='scatter',x='petal_width',y='petal_length',
                                     color='lightcoral',label='setosa')
data[data.species=='versicolor'].plot(kind='scatter',x='petal_width',y='petal_length',
                                     color='olive',label='versicolor',ax=sc)
data[data.species=='virginica'].plot(kind='scatter',x='petal_width',y='petal_length',
                                     color='lightseagreen', label='virginica', ax=sc)
sc.set_xlabel('petal width in cm',fontsize=15)
sc.set_ylabel('petal length in cm',fontsize=15)
sc.set_title('Petal Width vs Petal Length',fontsize=20)
sc=plt.gcf()
sc.set_size_inches(10,6)
```



9. Plot a Seaborn pairplot

```
sns.pairplot(data, hue='species', markers=["o", "s", "D"],palette='rocket')  
plt.show()
```

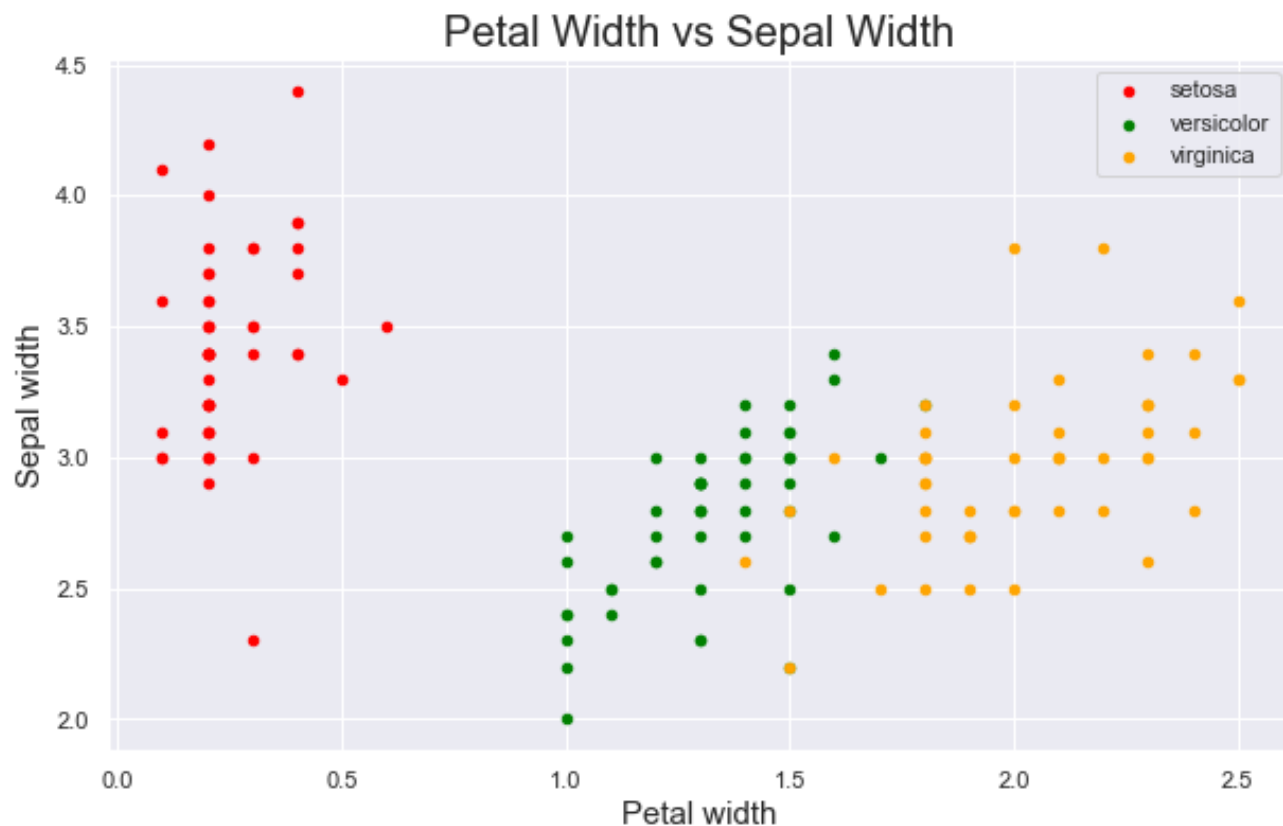
Pairplot of Iris Data



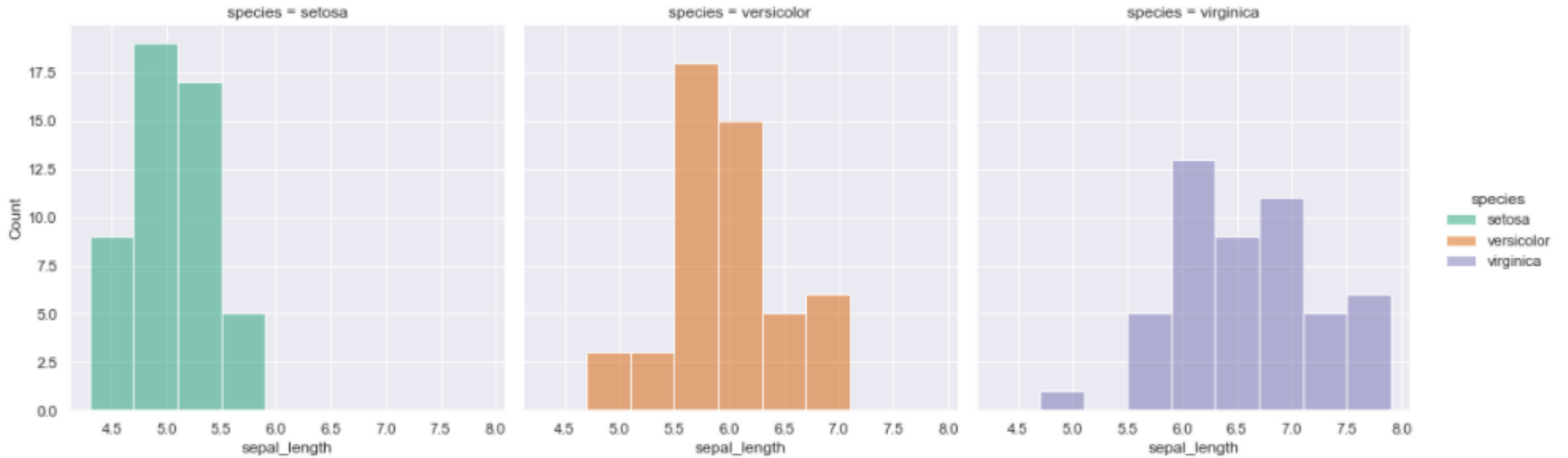
10. Plot any 2 plots of your choice from the Seaborn Library

1.

```
sns.set(style="darkgrid")
sc=data[data.species=='setosa'].plot(kind='scatter',x='petal_width',y='sepal_width',color='red',label='setosa')
data[data.species=='versicolor'].plot(kind='scatter',x='petal_width',y='sepal_width',color='green',label='versicolor',ax=sc)
data[data.species=='virginica'].plot(kind='scatter',x='petal_width',y='sepal_width',color='orange', label='virginica', ax=sc)
sc.set_xlabel('Petal width',fontsize=15)
sc.set_ylabel('Sepal width',fontsize=15)
sc.set_title('Petal Width vs Sepal Width',fontsize=20)
sc=plt.gcf()
sc.set_size_inches(10,6)
```



```
2. sns.displot(data=data, x="sepal_length", hue="species", col="species",palette='Dark2')  
plt.show()
```



The insights of the above visualisation:

1. Setosa has the highest sepal_length.
2. Virginica has the lowest sepal length.
3. The average count of sepal length all species of lies between 2.5 to 7.5