

```
[4]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[6]: #to load the iris data set
df = pd.read_csv("Iris_Data.csv")
print(df.head(5))
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

```
[7]: #info about the iris data set
df.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   species         150 non-null   object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```
[8]: #to view the unique species in the iris data set
df['species'].unique()
```

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

```
[9]: #to count the values under each species
df['species'].value_counts()
```

```
[9]: species
Iris-setosa      50
Iris-versicolor  50
Iris-virginica   50
Name: count, dtype: int64
```

```
[10]: #to select a single column
df['sepal_length']
```

```
[10]: 0      5.1
      1      4.9
      2      4.7
      3      4.6
      4      5.0
      ...
     145     6.7
     146     6.3
     147     6.5
     148     6.2
     149     5.9
      Name: sepal_length, Length: 150, dtype: float64
```

```
[16]: #to select multiple columns
df[['sepal_length', 'petal_length', 'species']]
```

```
[16]:
```

	sepal_length	petal_length	species
0	5.1	1.4	Iris-setosa
1	4.9	1.4	Iris-setosa
2	4.7	1.3	Iris-setosa
3	4.6	1.5	Iris-setosa
4	5.0	1.4	Iris-setosa
...
145	6.7	5.2	Iris-virginica
146	6.3	5.0	Iris-virginica
147	6.5	5.2	Iris-virginica
148	6.2	5.4	Iris-virginica
149	5.9	5.1	Iris-virginica

150 rows × 3 columns

```
[17]: #creating new column
df['sepal_ratio'] = df['sepal_length'] / df['sepal_width']
df.head(6)
```

```
[17]:
```

	sepal_length	sepal_width	petal_length	petal_width	species	is_large_petal	sepal_ratio
0	5.1	3.5	1.4	0.2	Iris-setosa	False	1.457143
1	4.9	3.0	1.4	0.2	Iris-setosa	False	1.633333
2	4.7	3.2	1.3	0.2	Iris-setosa	False	1.468750
3	4.6	3.1	1.5	0.2	Iris-setosa	False	1.483871
4	5.0	3.6	1.4	0.2	Iris-setosa	False	1.388889
5	5.4	3.9	1.7	0.4	Iris-setosa	False	1.384615

```
[17]: #to filter the rows
df[df['species']=='Iris-virginica']
```

[17]:	sepal_length	sepal_width	petal_length	petal_width	species
100	6.3	3.3	6.0	2.5	Iris-virginica
101	5.8	2.7	5.1	1.9	Iris-virginica
102	7.1	3.0	5.9	2.1	Iris-virginica
103	6.3	2.9	5.6	1.8	Iris-virginica
104	6.5	3.0	5.8	2.2	Iris-virginica
105	7.6	3.0	6.6	2.1	Iris-virginica
106	4.9	2.5	4.5	1.7	Iris-virginica
107	7.3	2.9	6.3	1.8	Iris-virginica
108	6.7	2.5	5.8	1.8	Iris-virginica
109	7.2	3.6	6.1	2.5	Iris-virginica
110	6.5	3.2	5.1	2.0	Iris-virginica
111	6.4	2.7	5.3	1.9	Iris-virginica
112	6.8	3.0	5.5	2.1	Iris-virginica

```
[18]: #to filter the rows
df[df['sepal_length']>6.0]
```

[18]:	sepal_length	sepal_width	petal_length	petal_width	species
50	7.0	3.2	4.7	1.4	Iris-versicolor
51	6.4	3.2	4.5	1.5	Iris-versicolor
52	6.9	3.1	4.9	1.5	Iris-versicolor
54	6.5	2.8	4.6	1.5	Iris-versicolor
56	6.3	3.3	4.7	1.6	Iris-versicolor
...
144	6.7	3.3	5.7	2.5	Iris-virginica
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

```
[19]: #using loc -> label based (row index + column names)
df.loc[0:4,['sepal_width','petal_width','species']]
```

```
[19]:
```

	sepal_width	petal_width	species
0	3.5	0.2	Iris-setosa
1	3.0	0.2	Iris-setosa
2	3.2	0.2	Iris-setosa
3	3.1	0.2	Iris-setosa
4	3.6	0.2	Iris-setosa

```
[24]: #using iloc -> position based (row number + column number)
df.iloc[10:16,0:4]
```

```
[24]:
```

	sepal_length	sepal_width	petal_length	petal_width
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

```
[26]: #creating a new column
df['sepal_ratio']=df['sepal_length']/df['sepal_width']
df['sepal_ratio']
```

```
[26]:
```

0	1.457143
1	1.633333
2	1.468750
3	1.483871
4	1.388889
...	
145	2.233333
146	2.520000
147	2.166667
148	1.823529
149	1.966667

Name: sepal_ratio, Length: 150, dtype: float64

```
[38]: #drop the 1st row i.e., 0th  
df.drop(0,axis=0)
```

```
[38]:
```

	sepal length	sepal_width	petal_length	petal_width	species
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-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

149 rows × 5 columns

```
[39]: #drop the last row i.e., 150th  
df.drop(149, axis=0)
```

```
[39]:
```

	sepal length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...
144	6.7	3.3	5.7	2.5	Iris-virginica
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 rows × 5 columns

```
•[11]: #finding mean by grouping into species
df.groupby('species').mean()
```

```
[11]:
```

	sepal_length	sepal_width	petal_length	petal_width
species				
Iris-setosa	5.006	3.418	1.464	0.244
Iris-versicolor	5.936	2.770	4.260	1.326
Iris-virginica	6.588	2.974	5.552	2.026

```
•[13]: #finding mean, min, max by grouping into species
df.groupby('species').agg(['mean','min','max'])
```

```
[13]:
```

	sepal_length			sepal_width			petal_length			petal_width		
	mean	min	max	mean	min	max	mean	min	max	mean	min	max
species												
Iris-setosa	5.006	4.3	5.8	3.418	2.3	4.4	1.464	1.0	1.9	0.244	0.1	0.6
Iris-versicolor	5.936	4.9	7.0	2.770	2.0	3.4	4.260	3.0	5.1	1.326	1.0	1.8
Iris-virginica	6.588	4.9	7.9	2.974	2.2	3.8	5.552	4.5	6.9	2.026	1.4	2.5

```
[15]: #creating flag column
df['is_large_petal'] = df['petal_length']>4.5
df.sort_values(by='petal_length',ascending=False)
```

```
[15]:
```

	sepal_length	sepal_width	petal_length	petal_width	species	is_large_petal
118	7.7	2.6	6.9	2.3	Iris-virginica	True
122	7.7	2.8	6.7	2.0	Iris-virginica	True
117	7.7	3.8	6.7	2.2	Iris-virginica	True
105	7.6	3.0	6.6	2.1	Iris-virginica	True
131	7.9	3.8	6.4	2.0	Iris-virginica	True
...
36	5.5	3.5	1.3	0.2	Iris-setosa	False
35	5.0	3.2	1.2	0.2	Iris-setosa	False
14	5.8	4.0	1.2	0.2	Iris-setosa	False
13	4.3	3.0	1.1	0.1	Iris-setosa	False
22	4.6	3.6	1.0	0.2	Iris-setosa	False

150 rows × 6 columns

```
[20]: df.to_csv("Iris_Data_Modified.csv", index=False)
      new_df = pd.read_csv("Iris_Data_Modified.csv")
      new_df.head()
```

```
[20]:
```

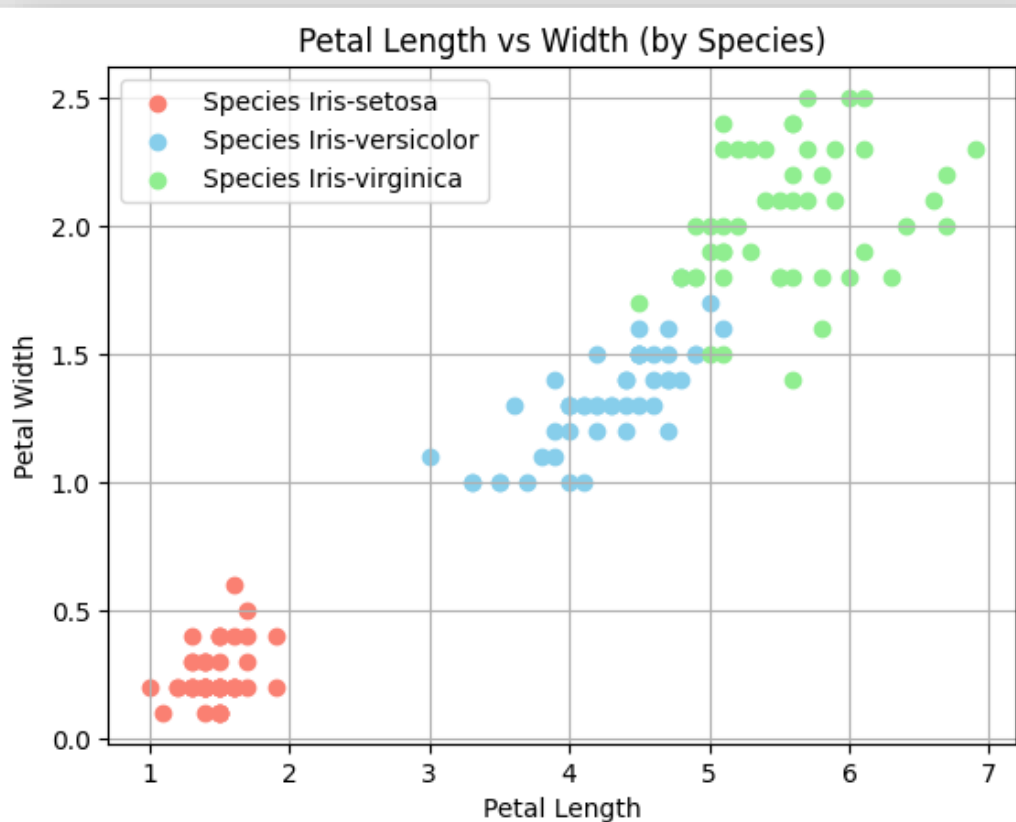
	sepal_length	sepal_width	petal_length	petal_width	species	is_large_petal	sepal_ratio
0	5.1	3.5	1.4	0.2	Iris-setosa	False	1.457143
1	4.9	3.0	1.4	0.2	Iris-setosa	False	1.633333
2	4.7	3.2	1.3	0.2	Iris-setosa	False	1.468750
3	4.6	3.1	1.5	0.2	Iris-setosa	False	1.483871
4	5.0	3.6	1.4	0.2	Iris-setosa	False	1.388889

```
[22]: #plot 1: Scatter plot : Sepal Length v/s Sepal Width
```

```
# Create a color map for species
colors = ['salmon', 'skyblue', 'lightgreen']
species = df['species'].unique()

# Plot
for i, s in enumerate(species):
    subset = df[df['species'] == s]
    plt.scatter(subset['petal_length'], subset['petal_width'], color=colors[i], label=f"Species {s}")

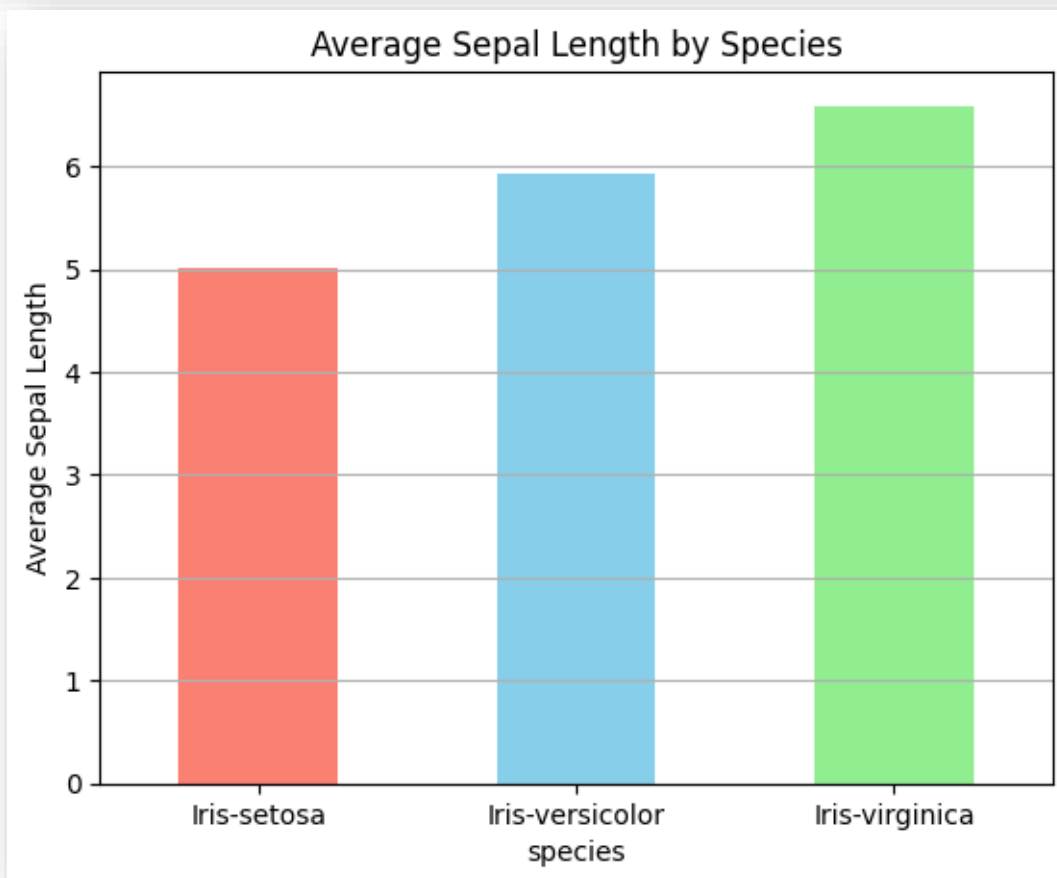
plt.xlabel("Petal Length")
plt.ylabel("Petal Width")
plt.title("Petal Length vs Width (by Species)")
plt.legend()
plt.grid(True)
plt.show()
```



[23]: *#plot 2: Bar chart - Average Sepal Length per Species*

```
# Group by species and get mean
mean_values = df.groupby('species')['sepal_length'].mean()

# Bar chart
mean_values.plot(kind='bar', color=['salmon', 'skyblue', 'lightgreen'])
plt.ylabel("Average Sepal Length")
plt.title("Average Sepal Length by Species")
plt.xticks(rotation=0)
plt.grid(axis='y')
plt.show()
```



[24]: *#plot 3: Violin Plot - Petal Length by Species*

```
sns.violinplot(x='species', y='petal_length', data=df, inner='box')
plt.title("Distribution of Petal Length by Species")
plt.xlabel("Species")
plt.ylabel("Petal Length")
plt.grid(True)
plt.show()
```