

```
In [1]: import pandas as pd
import matplotlib.pyplot as pt
import numpy as np
import requests as r
```

```
In [3]: df = pd.read_csv('iris.csv')
```

```
In [4]: df
```

Out[4]:

	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
...
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 × 5 columns

```
In [5]: df.columns=["sepal_length","sepal_width","petal_length","petal_width","species"]
df
```

Out[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
...
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 × 5 columns

In [6]:

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
```

In [7]:

df.isnull().sum()

```
Out[7]: sepal_length  0
sepal_width    0
petal_length    0
petal_width     0
species         0
dtype: int64
```

In [8]:

df.describe()

Out[8]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
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

In [9]:

df.drop_duplicates()

Out[9]:

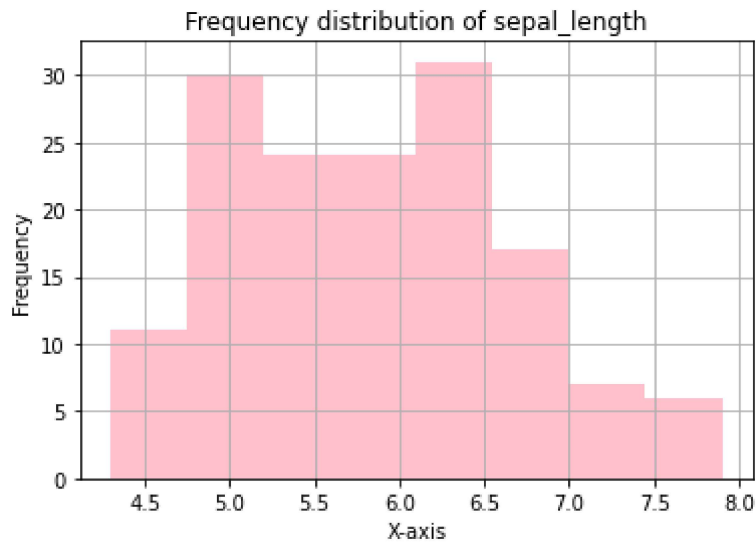
	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
...

	sepal_length	sepal_width	petal_length	petal_width	species
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

147 rows × 5 columns

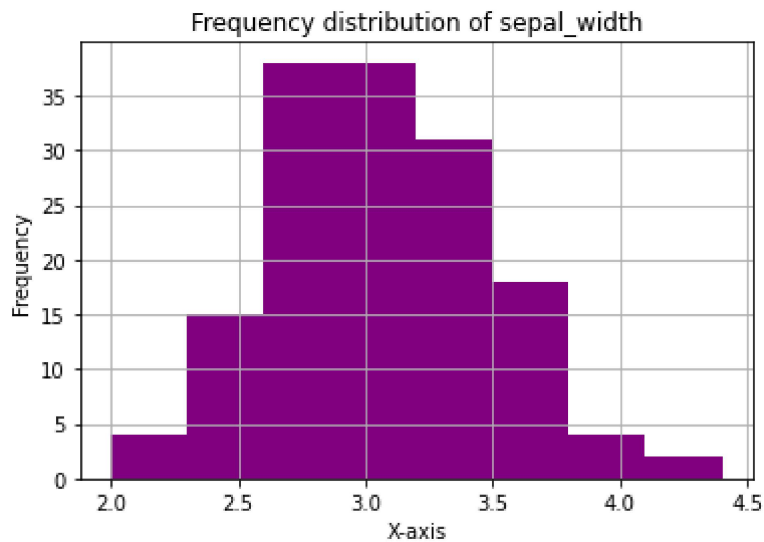
```
In [15]: pt.title("Frequency distribution of sepal_length")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["sepal_length"].hist(color="pink",bins=8)
```

```
Out[15]: <AxesSubplot:title={'center':'Frequency distribution of sepal_length'}, xlabel='X-axis', ylabel='Frequency'>
```



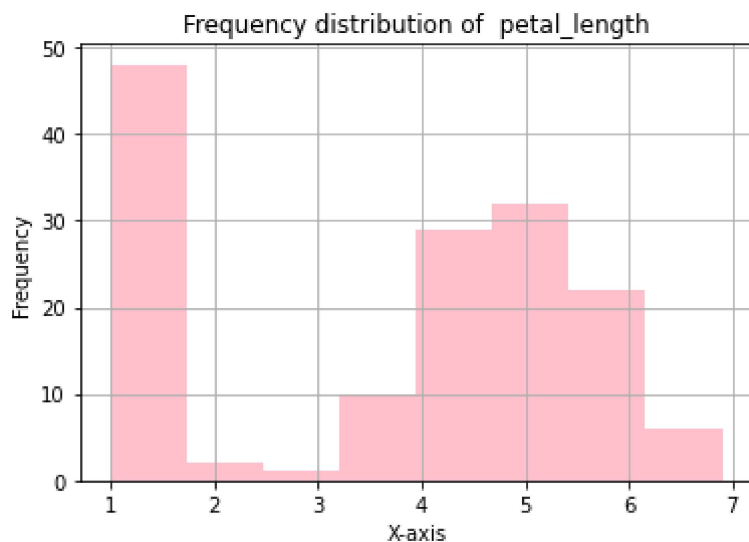
```
In [20]: pt.title("Frequency distribution of sepal_width")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["sepal_width"].hist(color="purple",bins=8)
```

```
Out[20]: <AxesSubplot:title={'center':'Frequency distribution of sepal_width'}, xlabel='X-axis', ylabel='Frequency'>
```



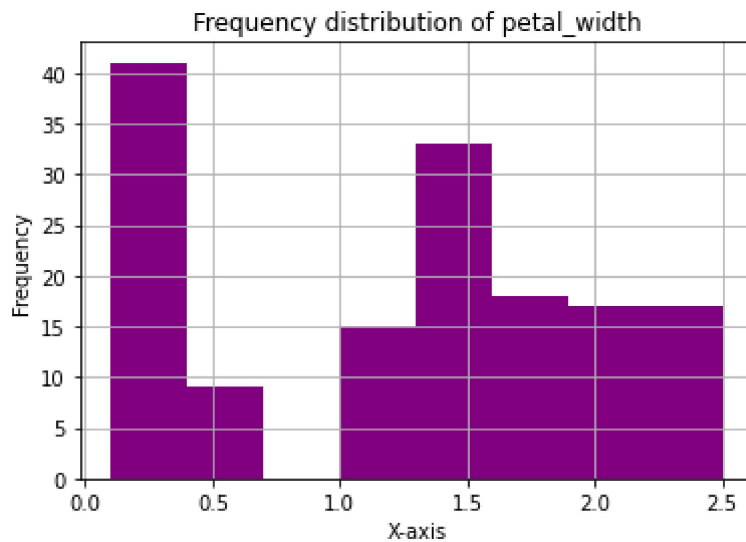
```
In [18]: pt.title("Frequency distribution of petal_length")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["petal_length"].hist(color="pink",bins=8)
```

```
Out[18]: <AxesSubplot:title={'center':'Frequency distribution of petal_length'}, xlabel='X-axis', ylabel='Frequency'>
```



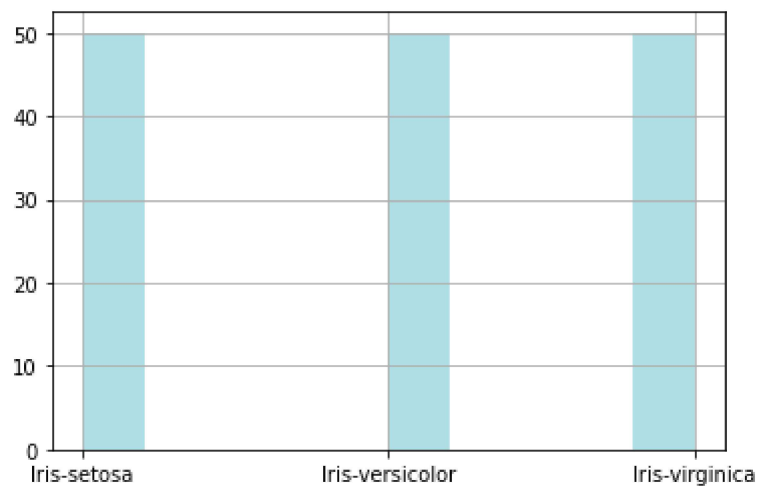
```
In [19]: pt.title("Frequency distribution of petal_width")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["petal_width"].hist(color="purple",bins=8)
```

```
Out[19]: <AxesSubplot:title={'center':'Frequency distribution of petal_width'}, xlabel='X-axis', ylabel='Frequency'>
```



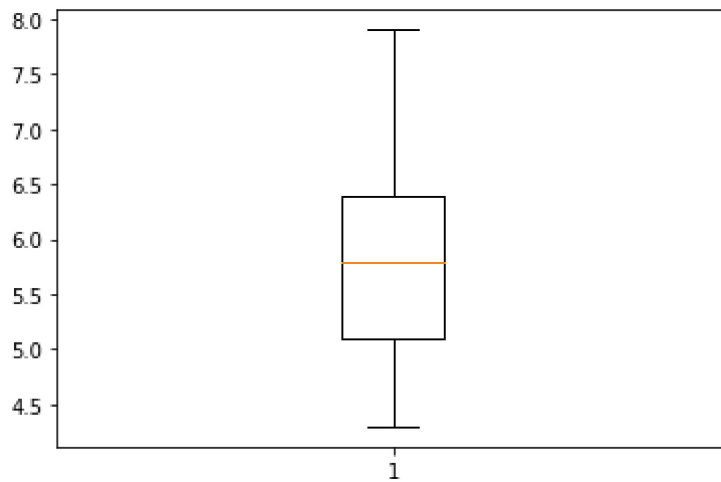
```
In [22]: df["species"].hist(color="powderblue")
```

Out[22]: <AxesSubplot:>



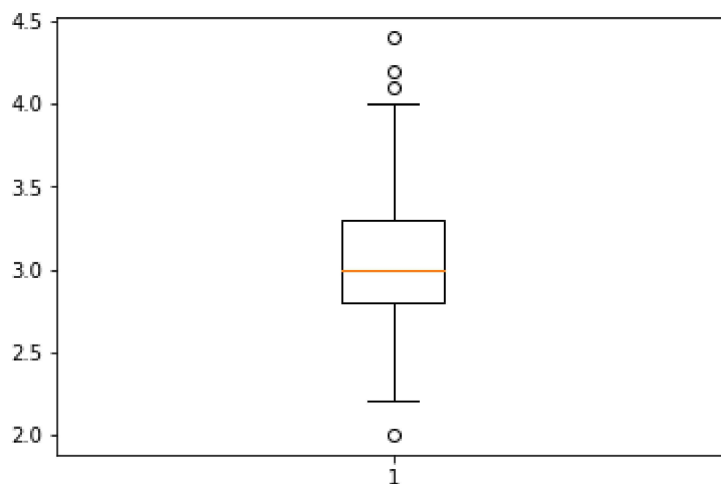
```
In [23]: pt.boxplot(df["sepal_length"])
```

Out[23]: {'whiskers': [<matplotlib.lines.Line2D at 0x1c335553a60>,
<matplotlib.lines.Line2D at 0x1c335553d30>],
'caps': [<matplotlib.lines.Line2D at 0x1c335563100>,
<matplotlib.lines.Line2D at 0x1c335563490>],
'boxes': [<matplotlib.lines.Line2D at 0x1c335553610>],
'medians': [<matplotlib.lines.Line2D at 0x1c335563820>],
'fliers': [<matplotlib.lines.Line2D at 0x1c335563bb0>],
'means': []}



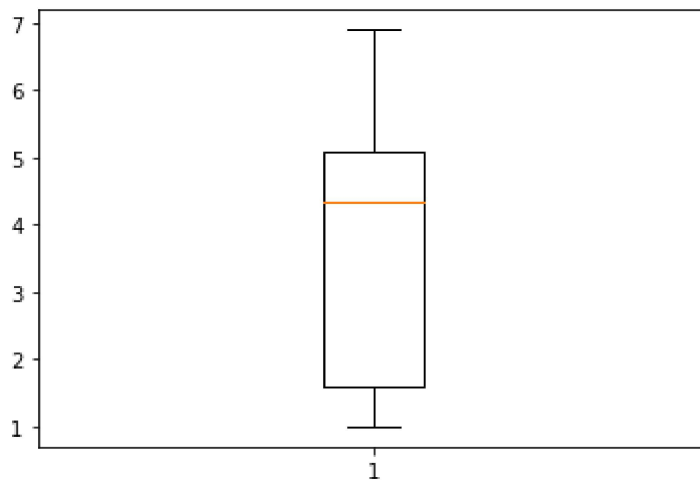
```
In [24]: pt.boxplot(df["sepal_width"])
```

```
Out[24]: {'whiskers': [<matplotlib.lines.Line2D at 0x1c3355bfe20>,  
  <matplotlib.lines.Line2D at 0x1c3355ce1f0>],  
  'caps': [<matplotlib.lines.Line2D at 0x1c3355ce580>,  
  <matplotlib.lines.Line2D at 0x1c3355ce910>],  
  'boxes': [<matplotlib.lines.Line2D at 0x1c3355bfa90>],  
  'medians': [<matplotlib.lines.Line2D at 0x1c3355ceca0>],  
  'fliers': [<matplotlib.lines.Line2D at 0x1c3355db070>],  
  'means': []}
```



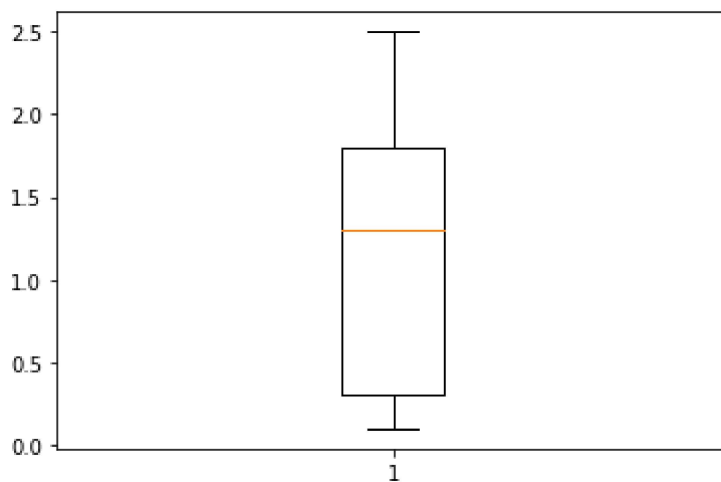
```
In [25]: pt.boxplot(df["petal_length"])
```

```
Out[25]: {'whiskers': [<matplotlib.lines.Line2D at 0x1c33562e7c0>,  
  <matplotlib.lines.Line2D at 0x1c33562eb50>],  
  'caps': [<matplotlib.lines.Line2D at 0x1c33562eee0>,  
  <matplotlib.lines.Line2D at 0x1c33563a2b0>],  
  'boxes': [<matplotlib.lines.Line2D at 0x1c33562e430>],  
  'medians': [<matplotlib.lines.Line2D at 0x1c33563a640>],  
  'fliers': [<matplotlib.lines.Line2D at 0x1c33563a9d0>],  
  'means': []}
```



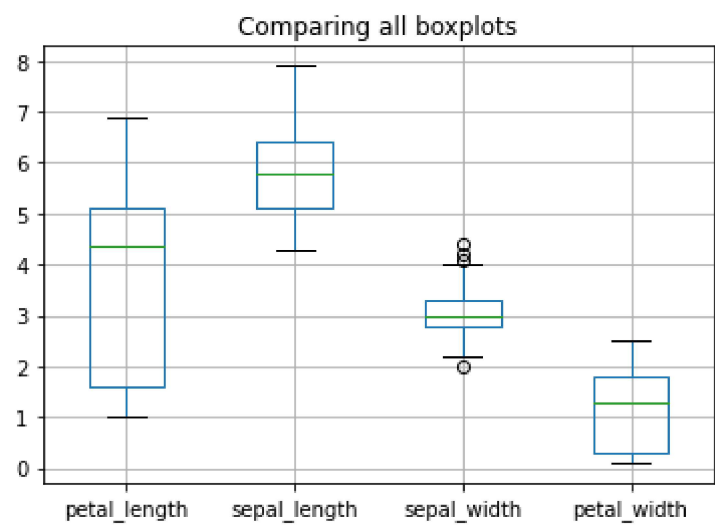
```
In [26]: pt.boxplot(df["petal_width"])
```

```
Out[26]: {'whiskers': [<matplotlib.lines.Line2D at 0x1c33569b6d0>,  
<matplotlib.lines.Line2D at 0x1c33569ba60>],  
'caps': [<matplotlib.lines.Line2D at 0x1c33569bdf0>,  
<matplotlib.lines.Line2D at 0x1c3356a61c0>],  
'boxes': [<matplotlib.lines.Line2D at 0x1c33569b310>],  
'medians': [<matplotlib.lines.Line2D at 0x1c3356a6550>],  
'fliers': [<matplotlib.lines.Line2D at 0x1c3356a68e0>],  
'means': []}
```



```
In [27]: df[["sepal_length", "sepal_width", "petal_length", "petal_width"]].boxplot()  
pt.title("Comparing all boxplots")
```

```
Out[27]: Text(0.5, 1.0, 'Comparing all boxplots')
```



```
In [ ]:
```