

```
In [2]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import requests as r
```

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

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

```
Out[4]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

```
In [10]: df.columns=["Id","SepalLengthCm","SepalWidthCm","PetalLengthCm","PetalWidthCm","Spe
```

```
In [11]: df
```

Out[11]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

In [12]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 6 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Id               150 non-null    int64
1   SepalLengthCm    150 non-null    float64
2   SepalWidthCm     150 non-null    float64
3   PetalLengthCm    150 non-null    float64
4   PetalWidthCm     150 non-null    float64
5   Species          150 non-null    object
dtypes: float64(4), int64(1), object(1)
memory usage: 7.2+ KB
```

In [13]: `df.isnull().sum()`

Out[13]:

Id	0
SepalLengthCm	0
SepalWidthCm	0
PetalLengthCm	0
PetalWidthCm	0
Species	0
dtype:	int64

In [14]: `df.describe()`

Out[14]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
count	150.000000	150.000000	150.000000	150.000000	150.000000
mean	75.500000	5.843333	3.054000	3.758667	1.198667
std	43.445368	0.828066	0.433594	1.764420	0.763161
min	1.000000	4.300000	2.000000	1.000000	0.100000
25%	38.250000	5.100000	2.800000	1.600000	0.300000
50%	75.500000	5.800000	3.000000	4.350000	1.300000
75%	112.750000	6.400000	3.300000	5.100000	1.800000
max	150.000000	7.900000	4.400000	6.900000	2.500000

In [15]: `df.drop_duplicates()`

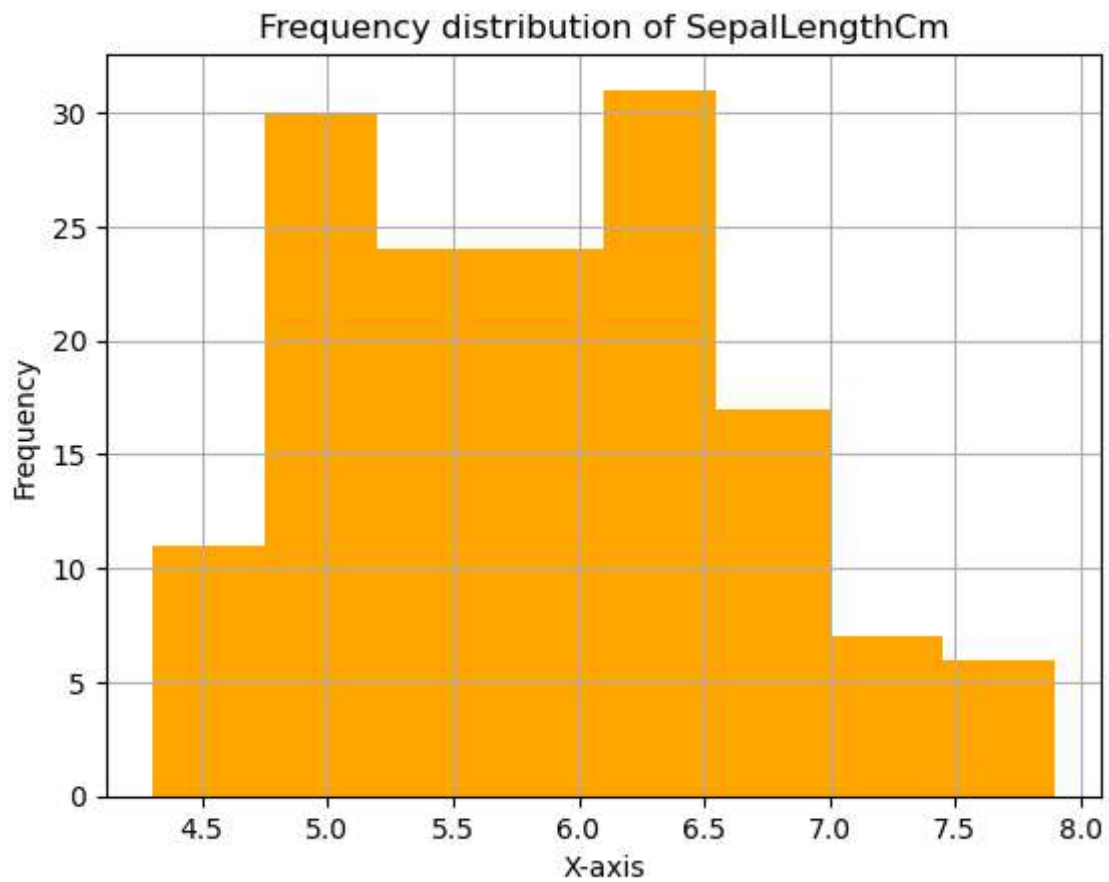
Out[15]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

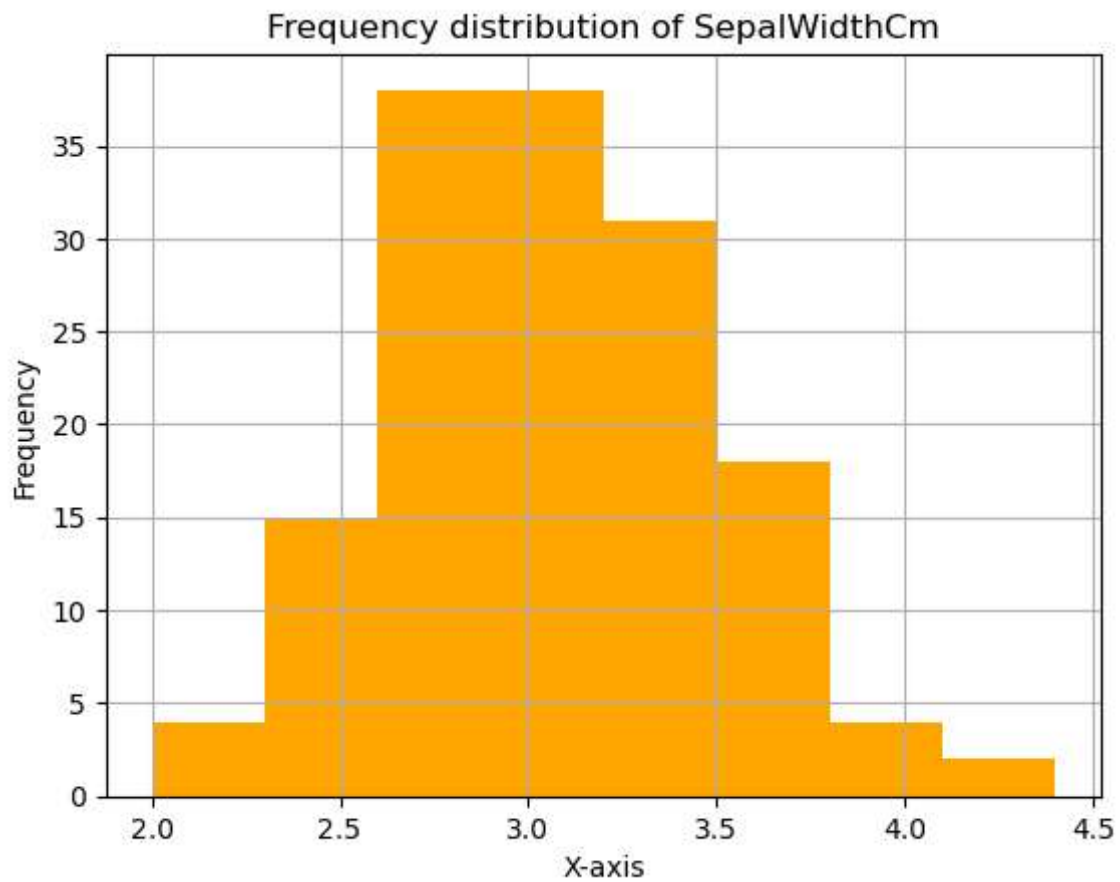
In [17]: `pt.title("Frequency distribution of SepalLengthCm")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["SepalLengthCm"].hist(color="orange",bins=8)`

Out[17]: `<Axes: title={'center': 'Frequency distribution of SepalLengthCm'}, xlabel='X-axis', ylabel='Frequency'>`



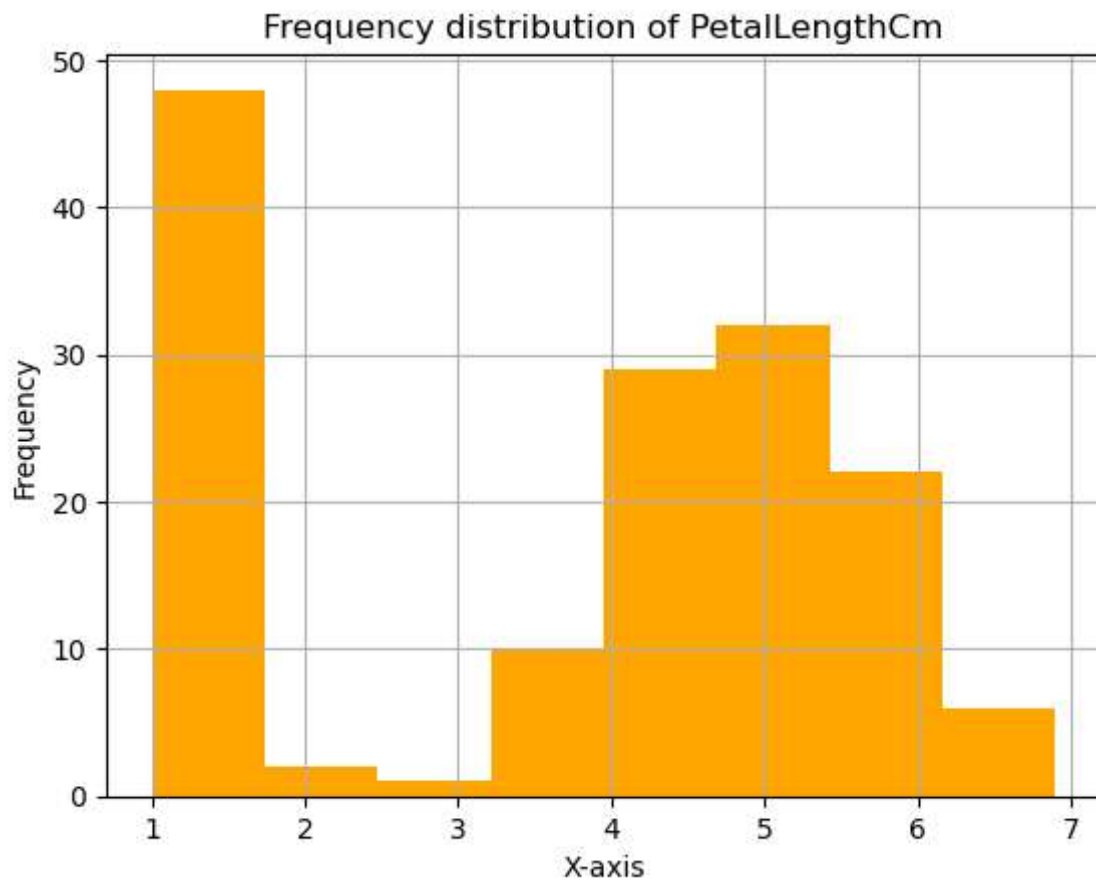
```
In [18]: pt.title("Frequency distribution of SepalWidthCm")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["SepalWidthCm"].hist(color="orange",bins=8)
```

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



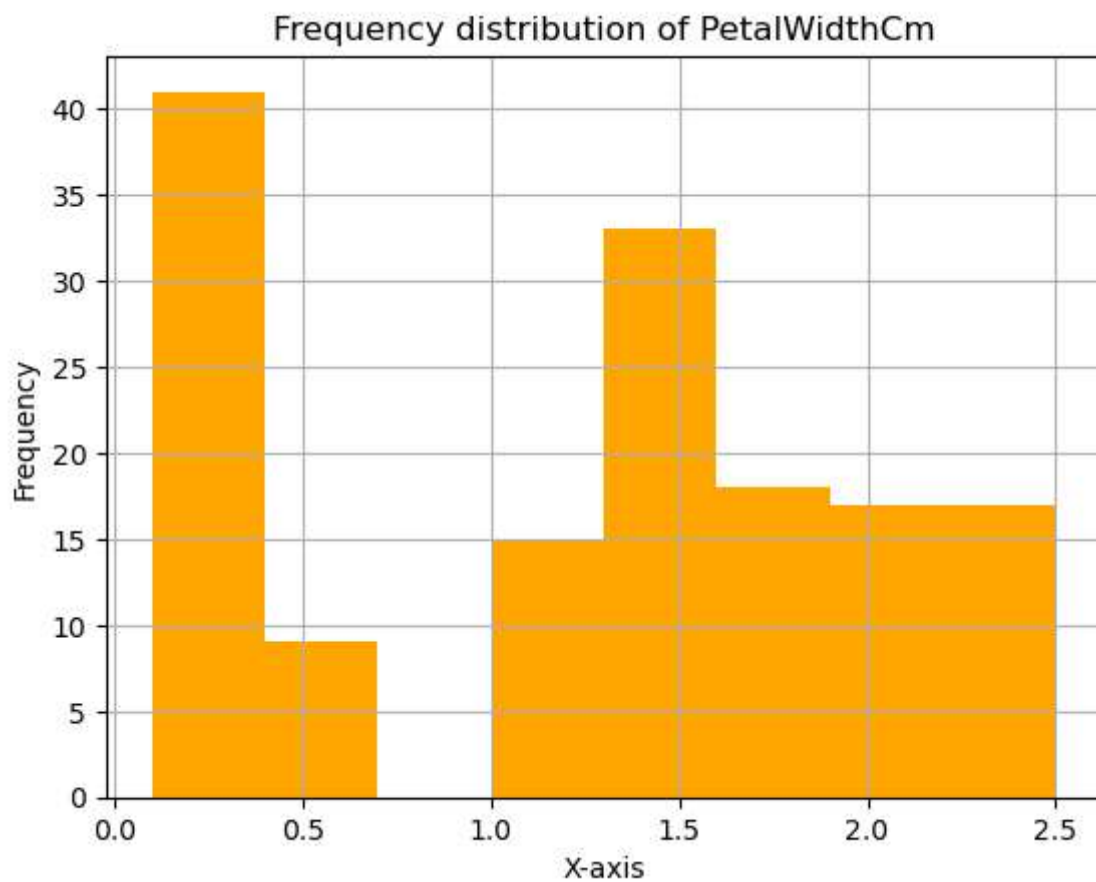
```
In [19]: pt.title("Frequency distribution of PetalLengthCm")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["PetalLengthCm"].hist(color="orange",bins=8)
```

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



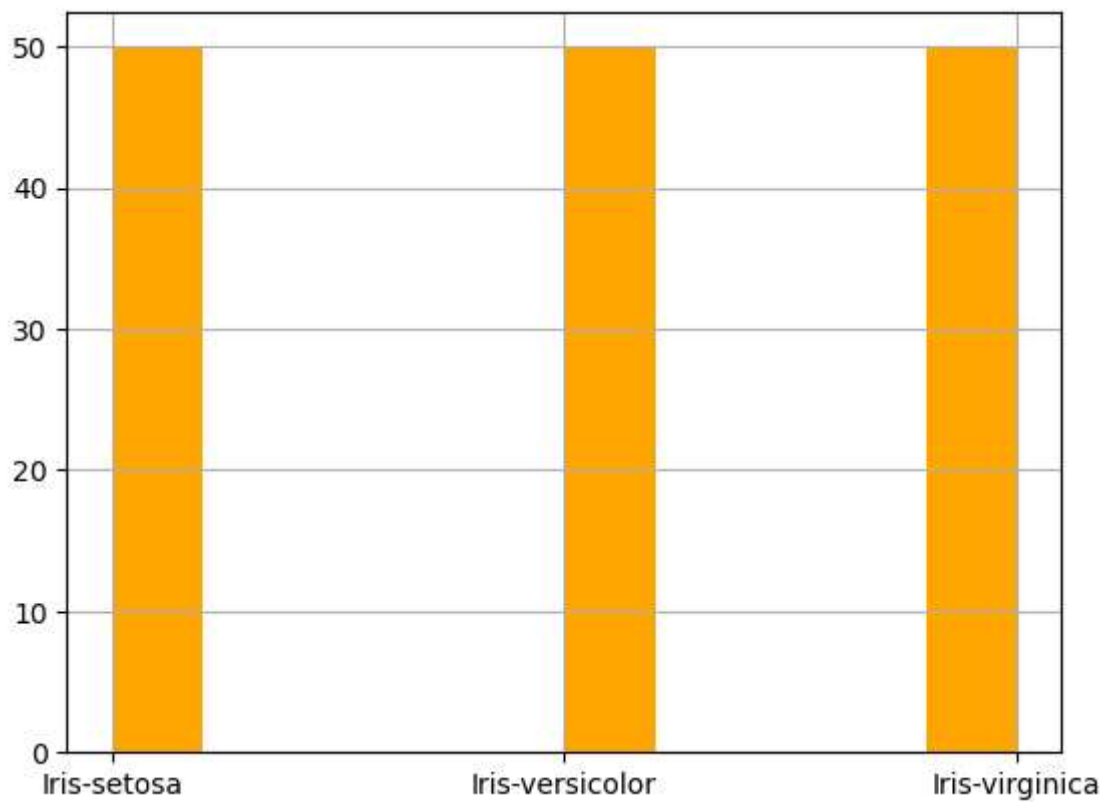
```
In [21]: pt.title("Frequency distribution of PetalWidthCm")
pt.xlabel("X-axis")
pt.ylabel("Frequency")
df["PetalWidthCm"].hist(color="orange",bins=8)
```

```
Out[21]: <Axes: title={'center': 'Frequency distribution of PetalWidthCm'}, xlabel='X-axis', ylabel='Frequency'>
```



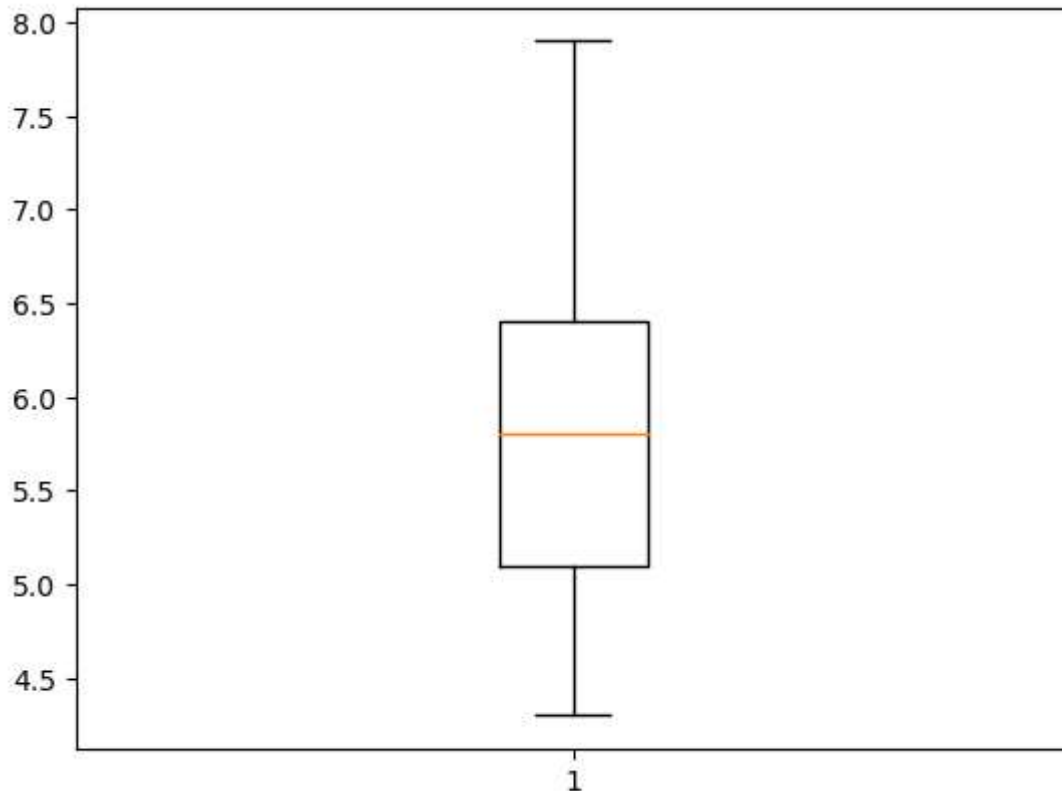
```
In [22]: df["Species"].hist(color="orange")
```

```
Out[22]: <Axes: >
```



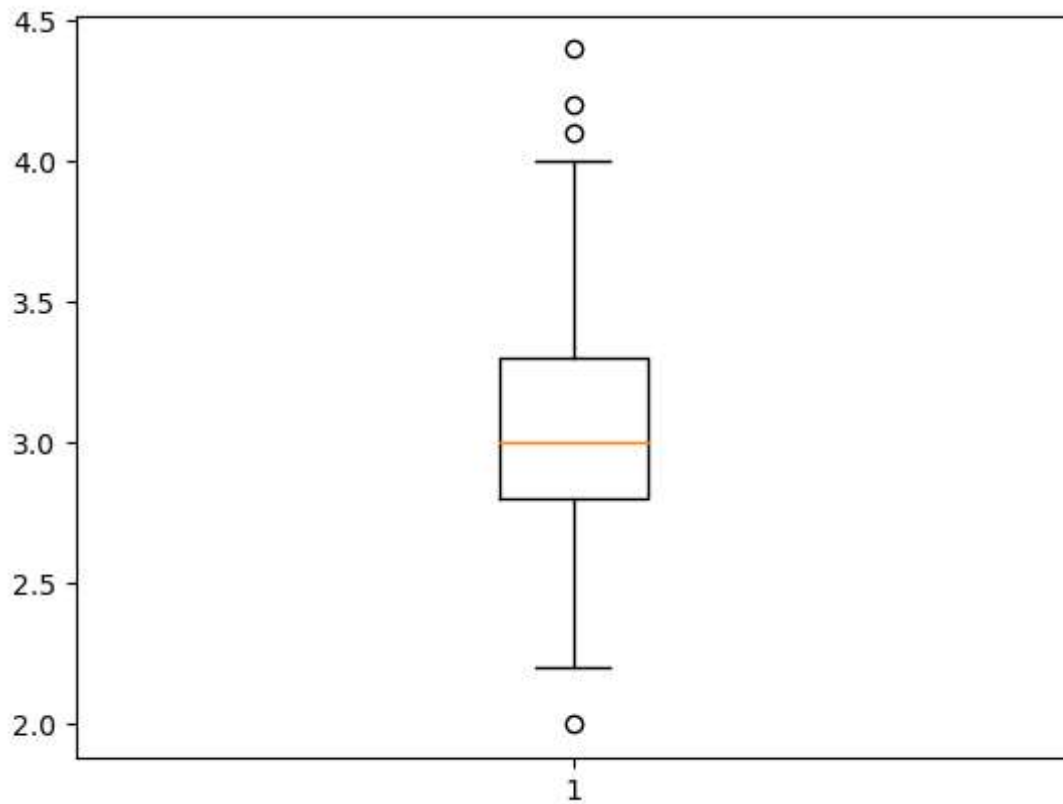
```
In [23]: pt.boxplot(df["SepalLengthCm"])
```

```
Out[23]: {'whiskers': [<matplotlib.lines.Line2D at 0x1555b9dbdc0>,
<matplotlib.lines.Line2D at 0x1555ba080a0>],
'caps': [<matplotlib.lines.Line2D at 0x1555ba08340>,
<matplotlib.lines.Line2D at 0x1555ba085e0>],
'boxes': [<matplotlib.lines.Line2D at 0x1555b9dbb20>],
'medians': [<matplotlib.lines.Line2D at 0x1555ba08880>],
'fliers': [<matplotlib.lines.Line2D at 0x1555ba08b20>],
'means': []}
```



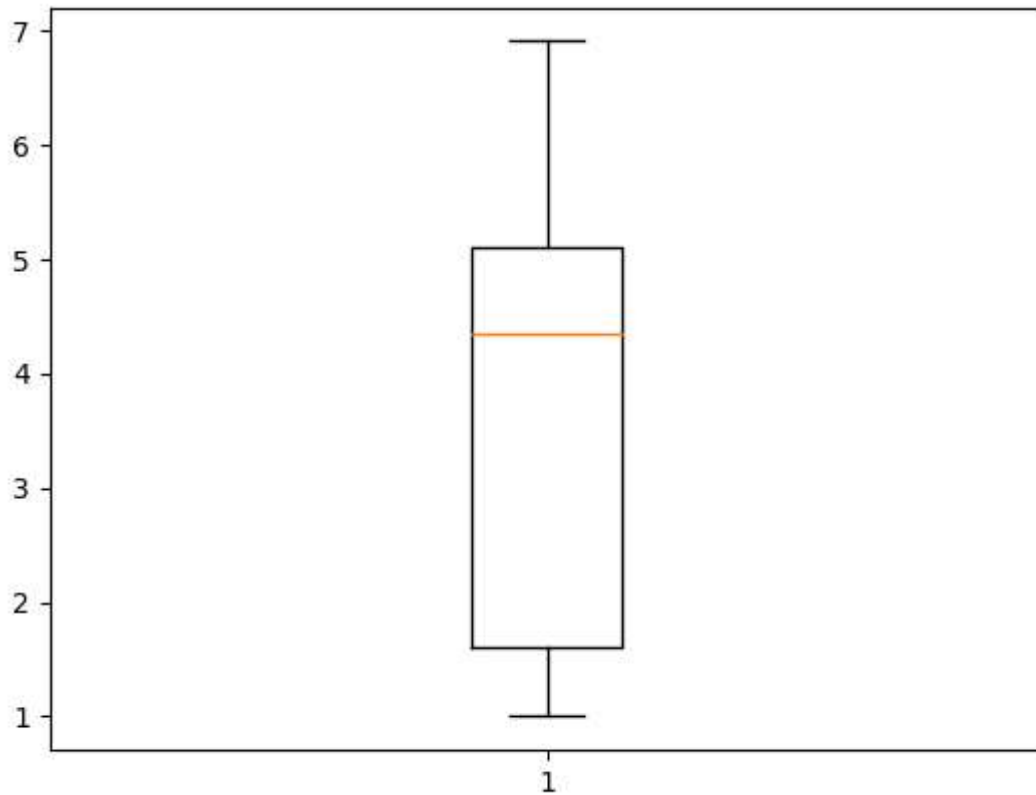
```
In [24]: pt.boxplot(df["SepalWidthCm"])
```

```
Out[24]: {'whiskers': [<matplotlib.lines.Line2D at 0x1555ca17a60>,
<matplotlib.lines.Line2D at 0x1555ca17d00>],
'caps': [<matplotlib.lines.Line2D at 0x1555ca17fa0>,
<matplotlib.lines.Line2D at 0x1555ca4c280>],
'boxes': [<matplotlib.lines.Line2D at 0x1555ca177c0>],
'medians': [<matplotlib.lines.Line2D at 0x1555ca4c520>],
'fliers': [<matplotlib.lines.Line2D at 0x1555ca4c7c0>],
'means': []}
```

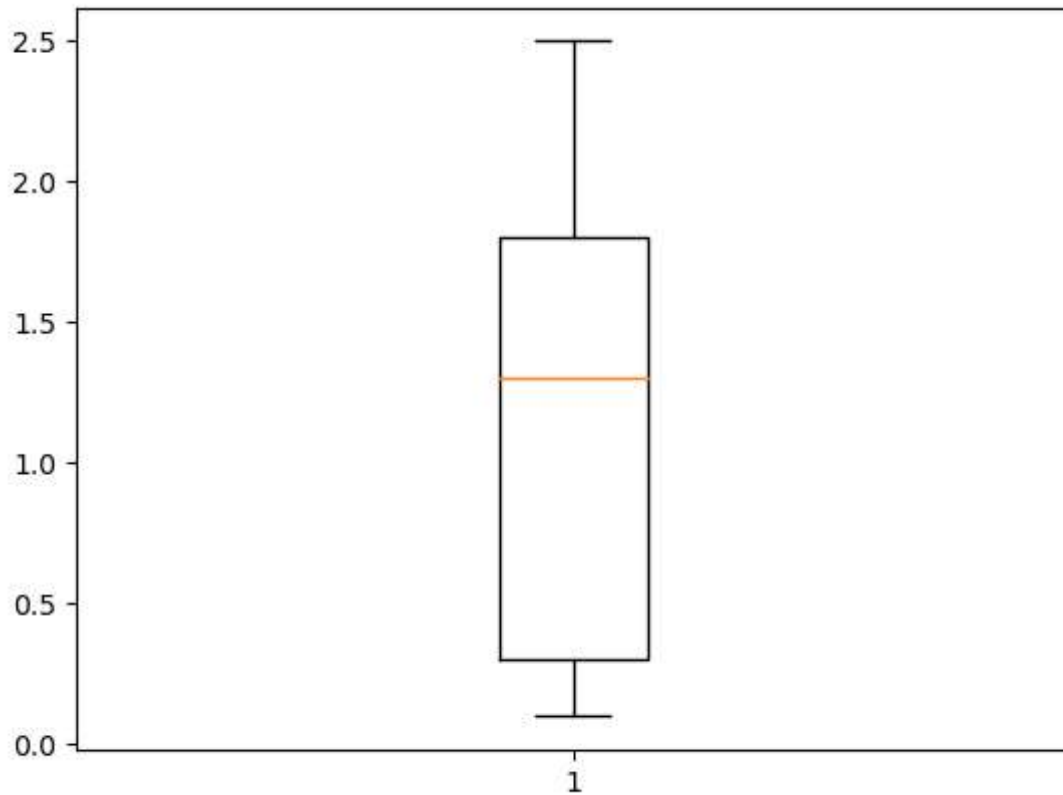
```
In [25]: pt.boxplot(df["PetalLengthCm"])
```

```
Out[25]: {'whiskers': [<matplotlib.lines.Line2D at 0x1555ca96950>,
<matplotlib.lines.Line2D at 0x1555ca96bf0>],
'caps': [<matplotlib.lines.Line2D at 0x1555ca96e90>,
<matplotlib.lines.Line2D at 0x1555ca97130>],
'boxes': [<matplotlib.lines.Line2D at 0x1555ca967d0>],
'medians': [<matplotlib.lines.Line2D at 0x1555ca973d0>],
'fliers': [<matplotlib.lines.Line2D at 0x1555ca97670>],
'means': []}
```



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

```
Out[26]: {'whiskers': [<matplotlib.lines.Line2D at 0x1555b6019f0>,
<matplotlib.lines.Line2D at 0x1555b601c90>],
'caps': [<matplotlib.lines.Line2D at 0x1555b601e10>,
<matplotlib.lines.Line2D at 0x1555b6020b0>],
'boxes': [<matplotlib.lines.Line2D at 0x1555b6017b0>],
'medians': [<matplotlib.lines.Line2D at 0x1555b602350>],
'fliers': [<matplotlib.lines.Line2D at 0x1555b6025f0>],
'means': []}
```

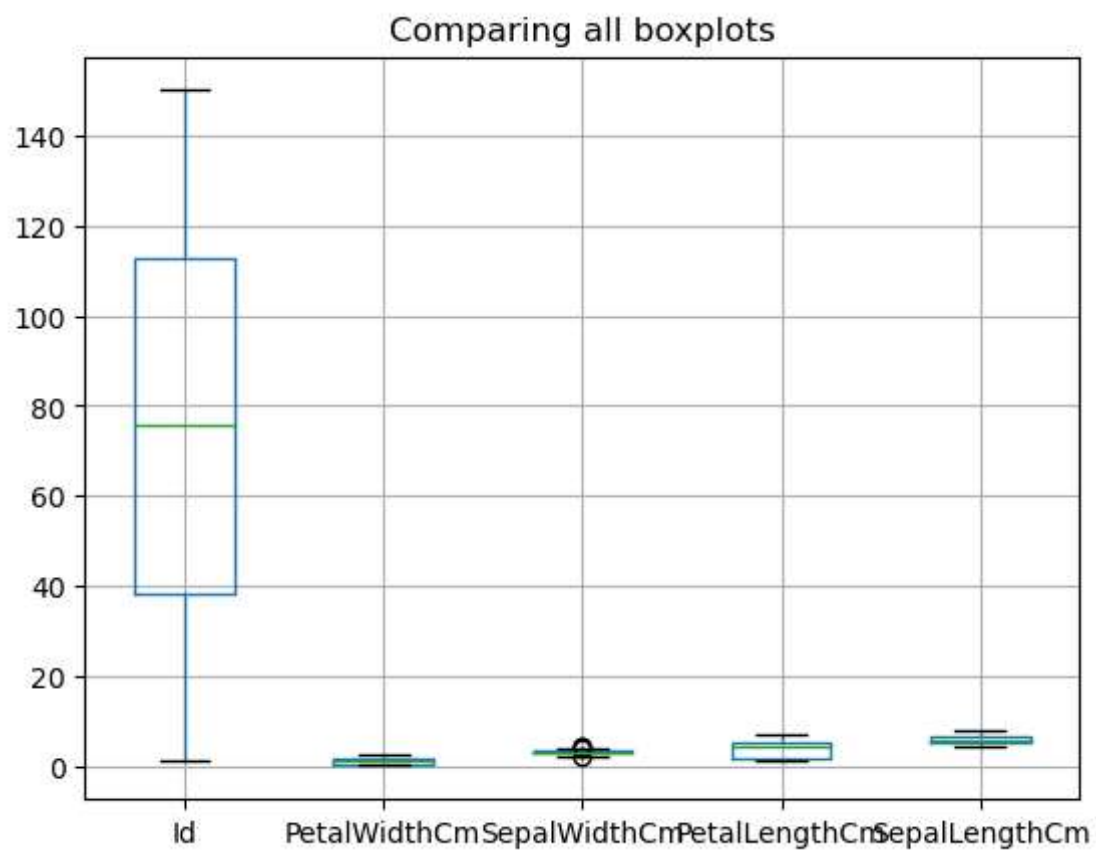


```
In [34]: df[{"Id", "SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", "Species"}].  
pt.title("Comparing all boxplots")
```

C:\Users\Aditi\AppData\Local\Temp\ipykernel_11120\2829034295.py:1: FutureWarning: Passing a set as an indexer is deprecated and will raise in a future version. Use a list instead.

```
df[{"Id", "SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", "Species"}].boxplot()
```

```
Out[34]: Text(0.5, 1.0, 'Comparing all boxplots')
```



In []: