ASSIGNMENT 10

Data Visualization 3

Download the Iris dataset and give the inference as :

1. List down features and their types
2. Create a histogram for each feature to illustrate feature distributions
3. Create a box plot for each feature
4. Compare distributions and identify outliers

> Importing required libraries and reading the dataset

In [1]:

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [2]:

```
df = pd.read_csv("Iris.csv")
```

In [3]:

```
df
```

Out[3]:

	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

> Data Preprocessing

In [4]:

```
#checks total size(rows*columns)
df.size
```

Out[4]:

```
900
```

In [5]:

```
#checks dimensions of dataframe
df.shape
```

Out[5]:

```
(150, 6)
```

In [6]:

```
#checks the columns present
df.columns
```

Out[6]:

```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWid
thCm',
       'Species'],
      dtype='object')
```

In [7]:

```
#prints information about the dataframe
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 [8]:

```
#checks initial statistics
df.describe()
```

Out[8]:

	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 [9]:

```
#checks datatype of each column
df.dtypes
```

Out[9]:

```
Id                 int64
SepalLengthCm    float64
SepalWidthCm     float64
PetalLengthCm    float64
PetalWidthCm     float64
Species           object
dtype: object
```

> Data Visualizations

The process of finding trends and correlations in our data by representing it pictorially is called Data Visualization.

DISTPLOT

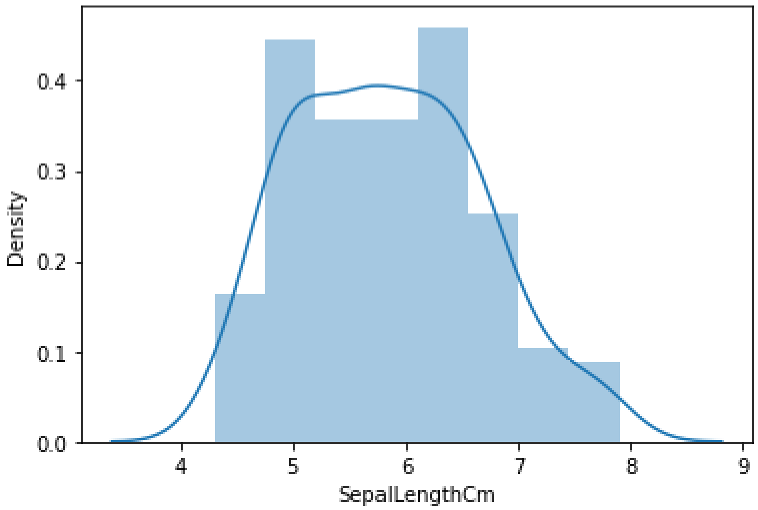
In [10]:

```
sns.distplot(df['SepalLengthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\distributi
ons.py:2551: FutureWarning: `distplot` is a deprecated function and will
be removed in a future version. Please adapt your code to use either `di
splot` (a figure-level function with similar flexibility) or `histplot`
(an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[10]:

```
<AxesSubplot:xlabel='SepalLengthCm', ylabel='Density'>
```



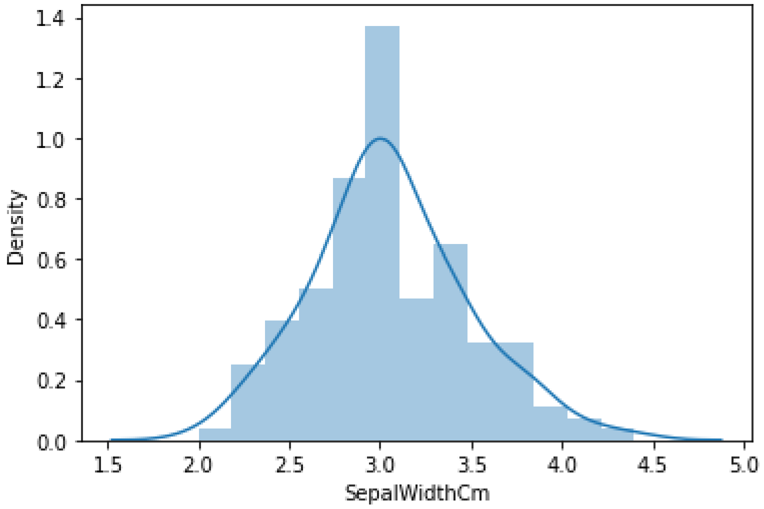
In [11]:

```
sns.distplot(df['SepalWidthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\distributi
ons.py:2551: FutureWarning: `distplot` is a deprecated function and will
be removed in a future version. Please adapt your code to use either `di
splot` (a figure-level function with similar flexibility) or `histplot`
(an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[11]:

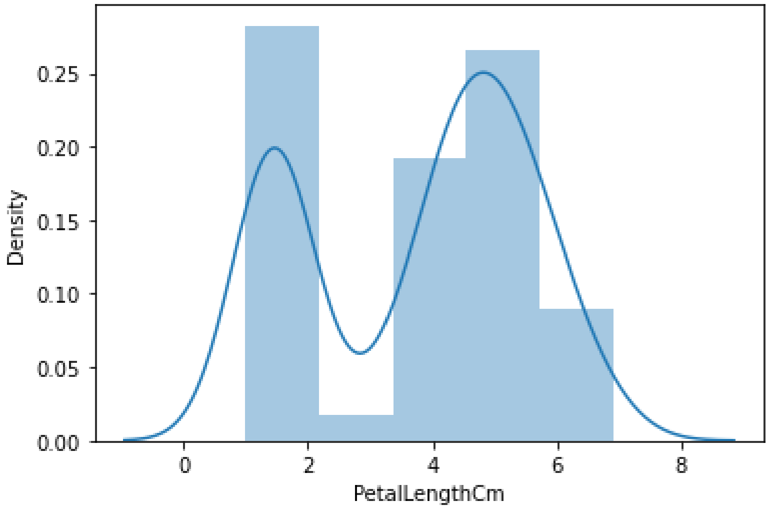
```
<AxesSubplot:xlabel='SepalWidthCm', ylabel='Density'>
```



```
In [12]:  sns.distplot(df['PetalLengthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\distributi
ons.py:2551: FutureWarning: `distplot` is a deprecated function and will
be removed in a future version. Please adapt your code to use either `di
splot` (a figure-level function with similar flexibility) or `histplot`
(an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

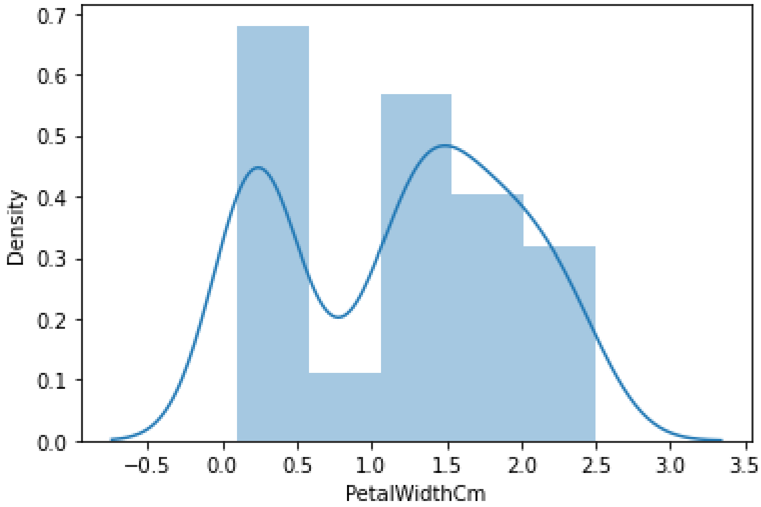
```
Out[12]: <AxesSubplot:xlabel='PetalLengthCm', ylabel='Density'>
```



```
In [13]:  sns.distplot(df['PetalWidthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\distributi
ons.py:2551: FutureWarning: `distplot` is a deprecated function and will
be removed in a future version. Please adapt your code to use either `di
splot` (a figure-level function with similar flexibility) or `histplot`
(an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

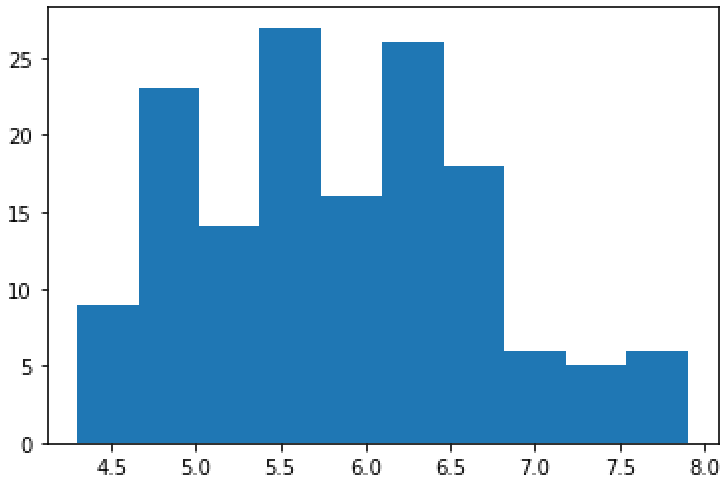
```
Out[13]: <AxesSubplot:xlabel='PetalWidthCm', ylabel='Density'>
```



HISTOGRAMS

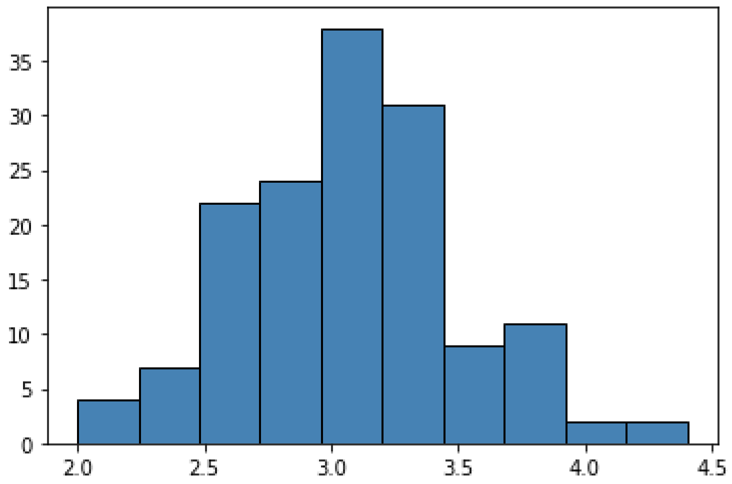
```
In [14]:   plt.hist(df['SepalLengthCm'])
```

```
Out[14]: (array([ 9., 23., 14., 27., 16., 26., 18.,  6.,  5.,  6.]),
          array([4.3 , 4.66, 5.02, 5.38, 5.74, 6.1 , 6.46, 6.82, 7.18, 7.54, 7.9
         ]),
          <BarContainer object of 10 artists>)
```



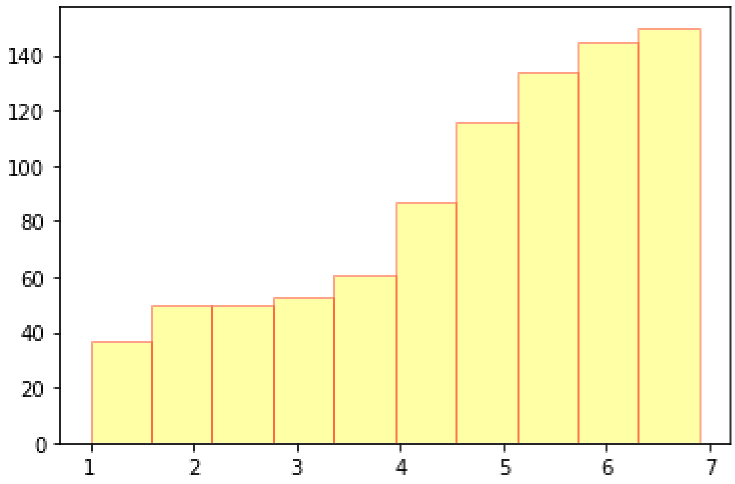
```
In [46]:   plt.hist(df['SepalWidthCm'],color='steelblue',edgecolor='black')
```

```
Out[46]: (array([ 4.,  7., 22., 24., 38., 31.,  9., 11.,  2.,  2.]),
          array([2.  , 2.24, 2.48, 2.72, 2.96, 3.2 , 3.44, 3.68, 3.92, 4.16, 4.4
         ]),
          <BarContainer object of 10 artists>)
```



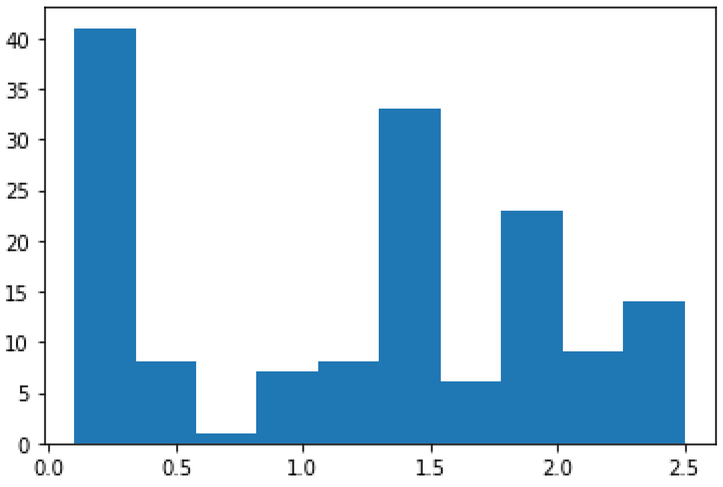
```
In [61]:  plt.hist(df['PetalLengthCm'],alpha=0.35,color='yellow',edgecolor='red',cur
```

```
Out[61]: (array([ 37.,  50.,  50.,  53.,  61.,  87., 116., 134., 145., 150.]),
          array([1.  , 1.59, 2.18, 2.77, 3.36, 3.95, 4.54, 5.13, 5.72, 6.31, 6.9
         ]),
          <BarContainer object of 10 artists>)
```



```
In [17]:  plt.hist(df['PetalWidthCm'])
```

```
Out[17]: (array([41.,  8.,  1.,  7.,  8., 33.,  6., 23.,  9., 14.]),
          array([0.1 , 0.34, 0.58, 0.82, 1.06, 1.3 , 1.54, 1.78, 2.02, 2.26, 2.5
         ]),
          <BarContainer object of 10 artists>)
```



BOXPLOTS

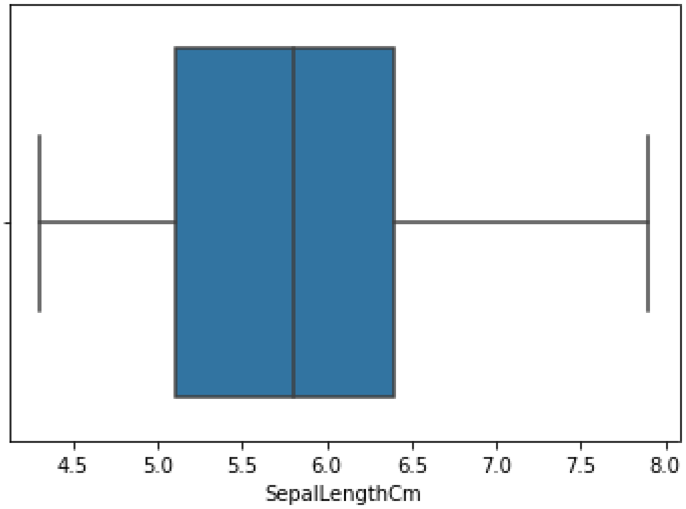
In [18]:

```
sns.boxplot(df.SepalLengthCm)
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variable as a keyword arg: x.
From version 0.12, the only valid positional argument will be `data`, an
d passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  warnings.warn(
```

Out[18]:

```
<AxesSubplot:xlabel='SepalLengthCm'>
```



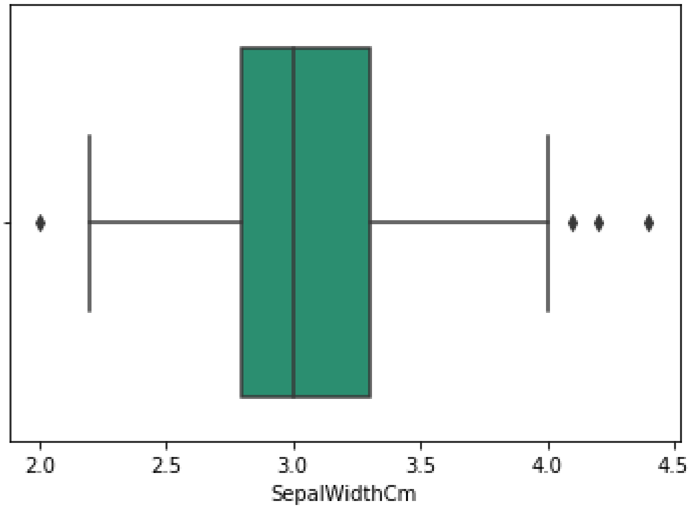
In [40]:

```
sns.boxplot(df.SepalWidthCm,palette='Dark2')
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variable as a keyword arg: x.
From version 0.12, the only valid positional argument will be `data`, an
d passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  warnings.warn(
```

Out[40]:

```
<AxesSubplot:xlabel='SepalWidthCm'>
```



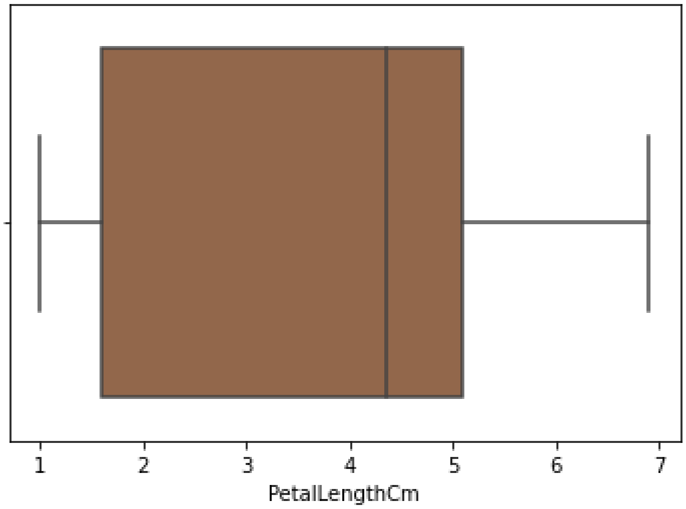
In [41]:

```
sns.boxplot(df.PetalLengthCm,palette='copper')
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variable as a keyword arg: x.
From version 0.12, the only valid positional argument will be `data`, an
d passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  warnings.warn(
```

Out[41]:

```
<AxesSubplot:xlabel='PetalLengthCm'>
```



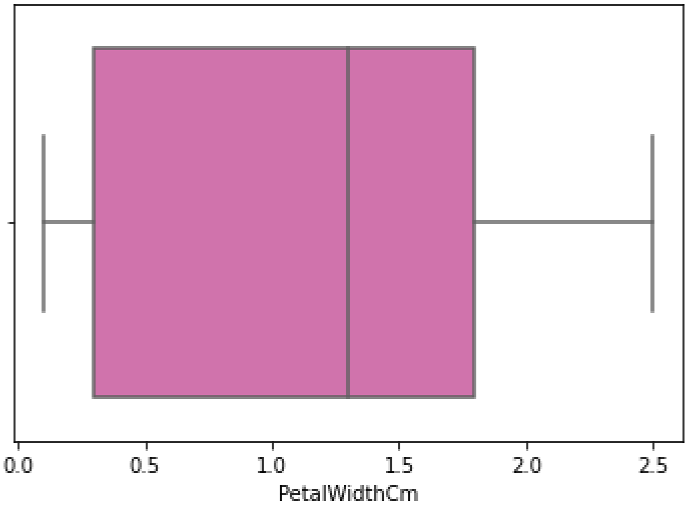
In [45]:

```
sns.boxplot(df.PetalWidthCm,palette='PuRd')
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variable as a keyword arg: x.
From version 0.12, the only valid positional argument will be `data`, an
d passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  warnings.warn(
```

Out[45]:

```
<AxesSubplot:xlabel='PetalWidthCm'>
```



DISPLOTS

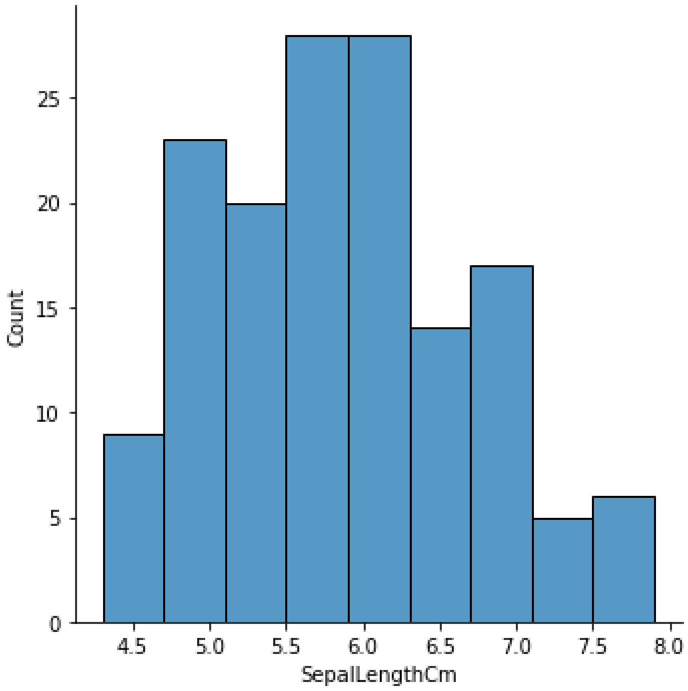
displot is a function in Seaborn used for creating various types of distribution plots. It provides a flexible interface to visualize univariate distributions of data. It can generate histograms, kernel density estimation (KDE) plots, rug plots, and more.

Here are the common parameters used in displot:

```
data: The dataset or DataFrame containing the data to be plotted.

x, y, hue: The variables from the data to be plotted on the x-axis, y-axis, or used for grouping, respectively.

kind: The type of distribution plot to be created. It can be "hist" for a histogram, "kde" for a kernel density plot, or "ecdf" for an empirical cumulative distribution function plot.

rug: If True, a rug plot will be displayed along the axis showing individual data points.

bins: The number of bins for the histogram, applicable when kind="hist".

kde: If True, a kernel density estimation plot will be overlaid on the histogram, applicable when kind="hist".

common_norm: If True, the distribution plots will be scaled to have a common y-axis scale.

height, aspect: The height and aspect ratio of the figure, respectively.

color: The color of the elements in the plot.

multiple: If True, multiple distribution plots will be created for different subsets of the data, based on the hue parameter.
```

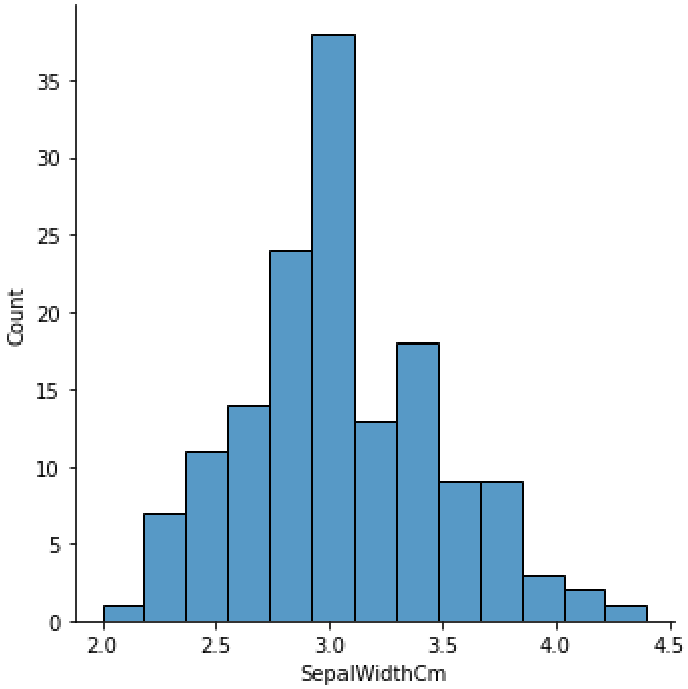
```
In [22]:  sns.displot(df['SepalLengthCm'])
```

```
Out[22]: <seaborn.axisgrid.FacetGrid at 0x1f0ad8d0790>
```



```
In [34]:  sns.displot(df['SepalWidthCm'],kind='hist')
```

```
Out[34]: <seaborn.axisgrid.FacetGrid at 0x1f0adab1fa0>
```

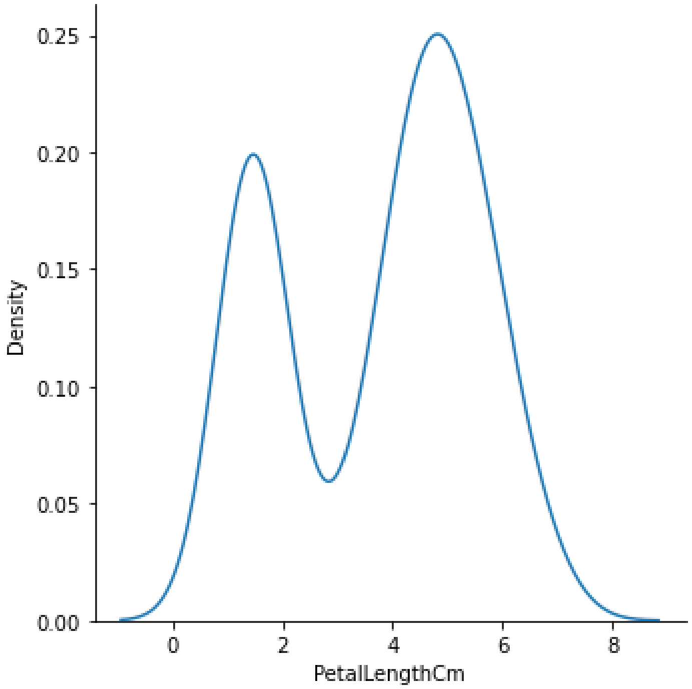


In [31]:

```
sns.displot(df['PetalLengthCm'],kind='kde')
```

Out[31]:

```
<seaborn.axisgrid.FacetGrid at 0x1f0aeced7c0>
```

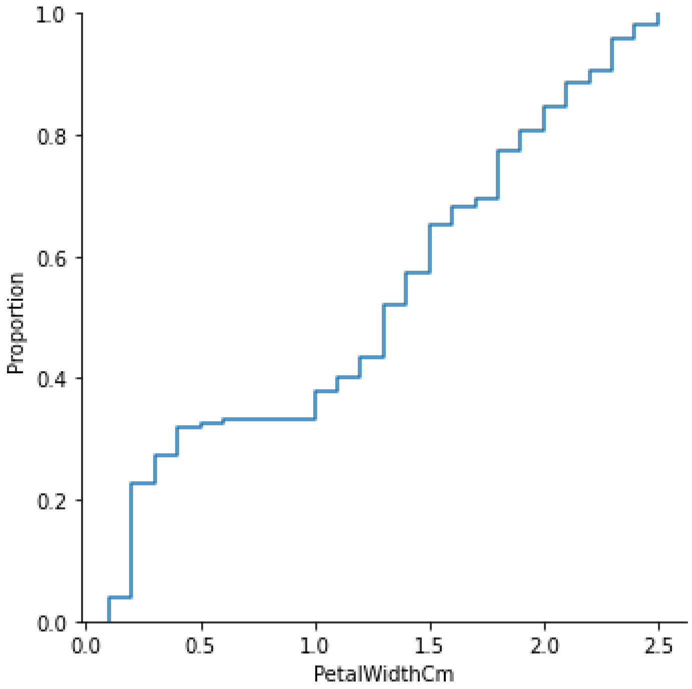


In [30]:

```
sns.displot(df['PetalWidthCm'],kind='ecdf')
```

Out[30]:

```
<seaborn.axisgrid.FacetGrid at 0x1f0aec96610>
```

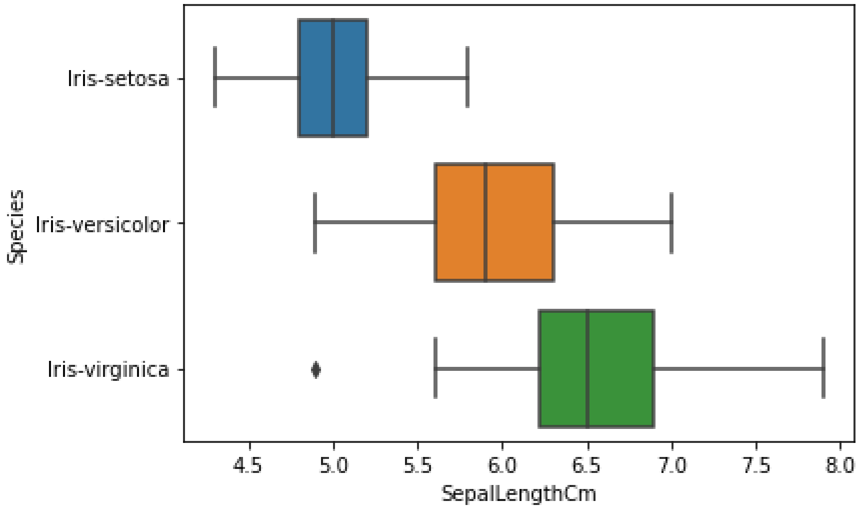


Comparisons

```
In [26]: sns.boxplot(df['SepalLengthCm'],df['Species'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variables as keyword args: x,
y. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in a
n error or misinterpretation.
  warnings.warn(
```

```
Out[26]: <AxesSubplot:xlabel='SepalLengthCm', ylabel='Species'>
```



Outliers present for Iris-virginica in SepalLengthCm.

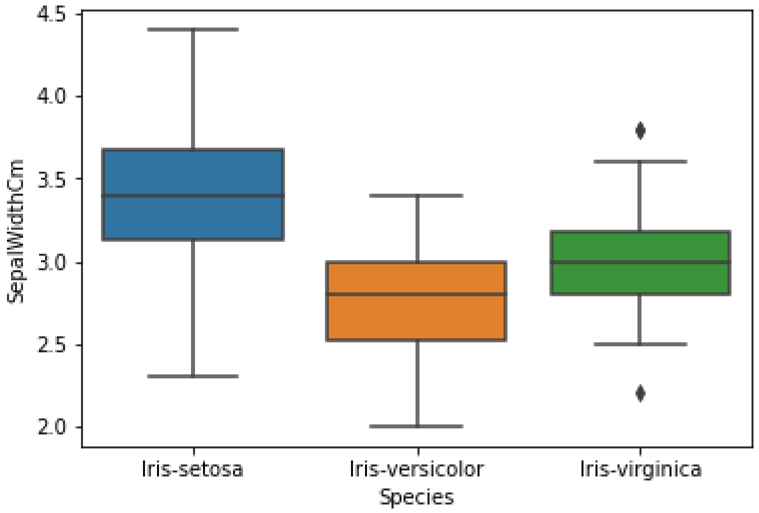
In [27]:

```
sns.boxplot(df['Species'],df['SepalWidthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variables as keyword args: x,
y. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in a
n error or misinterpretation.
  warnings.warn(
```

Out[27]:

```
<AxesSubplot:xlabel='Species', ylabel='SepalWidthCm'>
```



Outliers present for Iris-virginica in SepalWidthCm.

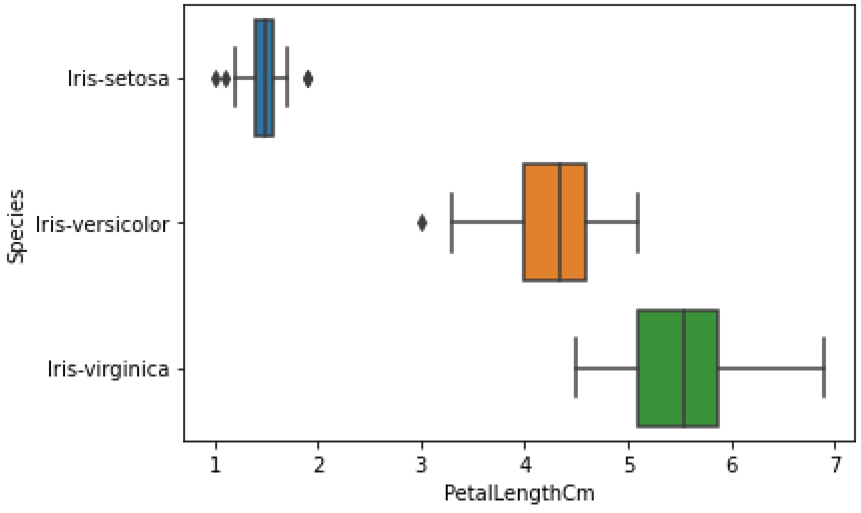
In [28]:

```
sns.boxplot(df['PetalLengthCm'],df['Species'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variables as keyword args: x,
y. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in a
n error or misinterpretation.
  warnings.warn(
```

Out[28]:

```
<AxesSubplot:xlabel='PetalLengthCm', ylabel='Species'>
```

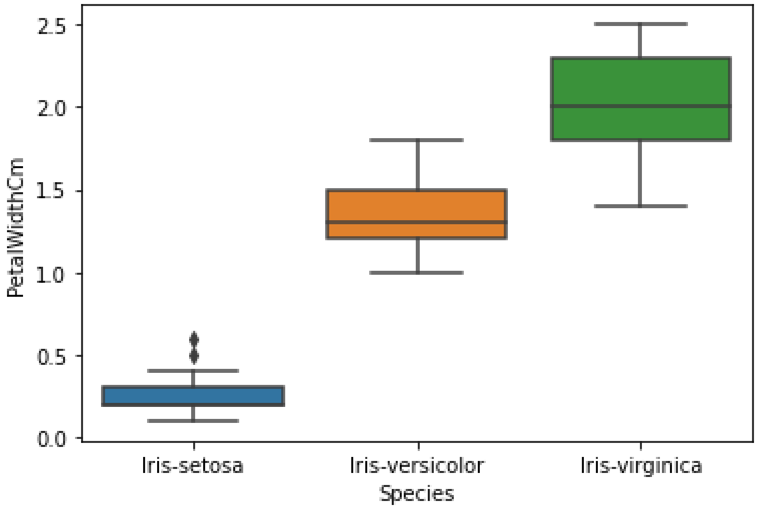


Outliers present for Iris-setosa and Iris-ersicolor in PetalLengthCm.

```
In [29]:   sns.boxplot(df['Species'],df['PetalWidthCm'])
```

```
C:\Users\Shravani Sajekar\anaconda3\lib\site-packages\seaborn\_decorator
s.py:36: FutureWarning: Pass the following variables as keyword args: x,
y. From version 0.12, the only valid positional argument will be `data`,
and passing other arguments without an explicit keyword will result in a
n error or misinterpretation.
  warnings.warn(
```

```
Out[29]: <AxesSubplot:xlabel='Species', ylabel='PetalWidthCm'>
```



Outliers present for Iris-setosa in PetalWidthCm.