In [ ]:
```
NAME ; ARYAN SIRDESAI
ROLL NO.: TACO20175
Lab Assignment 10 : Data Visualization III

Problem Statement : Download the Iris flower dataset or any other dataset into a Da
https://archive.ics.uci.edu/ml/datasets/Iris). Scan the dataset and give the infer
1. List down the features and their types (e.g., numeric, nominal) available in th
2. Create a histogram for each feature in the dataset to illustrate the feature di
3. Create a boxplot for each feature in the dataset.
4. Compare distributions and identify outliers.
```

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
iris = pd.read_csv('Iris.csv')
iris
```

Out[2]:

	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 [3]:
```
iris.head(20)
```

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
5	6	5.4	3.9	1.7	0.4	Iris-setosa
6	7	4.6	3.4	1.4	0.3	Iris-setosa
7	8	5.0	3.4	1.5	0.2	Iris-setosa
8	9	4.4	2.9	1.4	0.2	Iris-setosa
9	10	4.9	3.1	1.5	0.1	Iris-setosa
10	11	5.4	3.7	1.5	0.2	Iris-setosa
11	12	4.8	3.4	1.6	0.2	Iris-setosa
12	13	4.8	3.0	1.4	0.1	Iris-setosa
13	14	4.3	3.0	1.1	0.1	Iris-setosa
14	15	5.8	4.0	1.2	0.2	Iris-setosa
15	16	5.7	4.4	1.5	0.4	Iris-setosa
16	17	5.4	3.9	1.3	0.4	Iris-setosa
17	18	5.1	3.5	1.4	0.3	Iris-setosa
18	19	5.7	3.8	1.7	0.3	Iris-setosa
19	20	5.1	3.8	1.5	0.3	Iris-setosa

In [4]:
```
iris.describe()
```

Out[4]:

	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 [5]:
```
iris.isnull().sum()
```

Out[5]:
```
Id              0
SepalLengthCm   0
SepalWidthCm    0
PetalLengthCm   0
PetalWidthCm    0
Species         0
dtype: int64
```

In [6]:
```
iris.columns
```

Out[6]:
```
Index(['Id', 'SepalLengthCm', 'SepalWidthCm', 'PetalLengthCm', 'PetalWidthCm',
       'Species'],
      dtype='object')
```

SepalLengthCm : Numerical variable

SepalWidthCm: Numerical variable

PetalLengthCm: Numerical variable

PetalWidthCm: Numerical variable

Species: Categorial Variable

```
Categories: Iris-setosa
            Iris-versicolor
            Iris-virginica
```

In [7]:
```
iris.dtypes
```

Out[7]:
```
Id                int64
SepalLengthCm   float64
SepalWidthCm    float64
PetalLengthCm   float64
PetalWidthCm    float64
Species          object
dtype: object
```

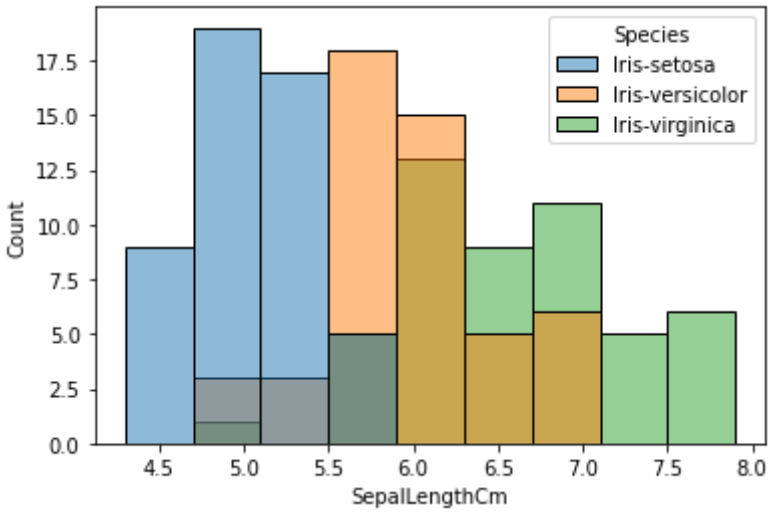
In [8]:
```
sns.countplot(iris['Species'])
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/_decorators.py:36: FutureWar
ning: Pass the following variable as a keyword arg: x. From version 0.12, the only
valid positional argument will be `data`, and passing other arguments without an e
xplicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[8]:
```
<AxesSubplot:xlabel='Species', ylabel='count'>
```

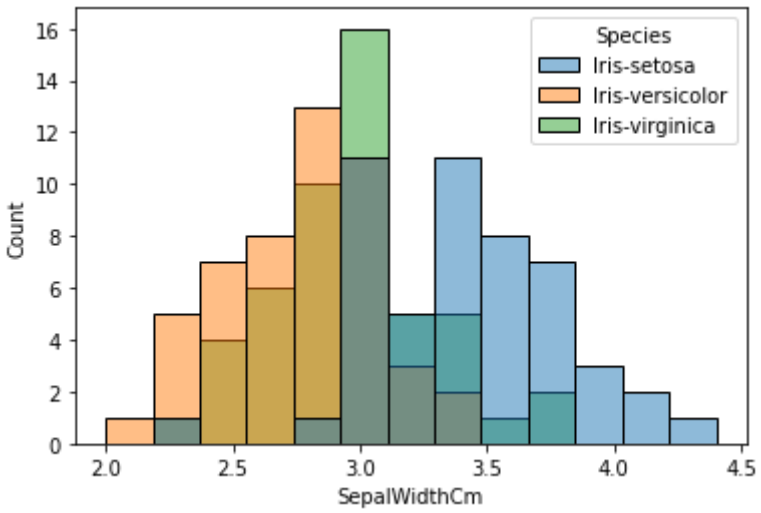
In [9]:
```
sns.histplot(data=iris, x="SepalLengthCm",hue="Species")
```

Out[9]:
```
<AxesSubplot:xlabel='SepalLengthCm', ylabel='Count'>
```



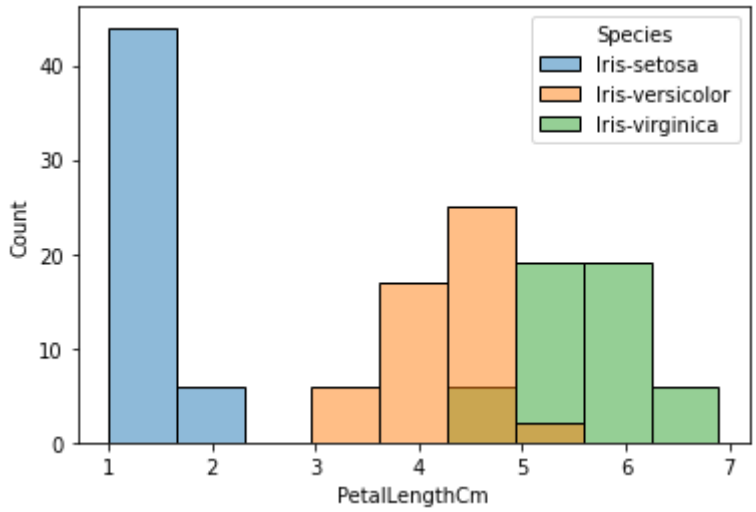
In [10]:
```
sns.histplot(data=iris, x="SepalWidthCm",hue="Species")
```

Out[10]:
```
<AxesSubplot:xlabel='SepalWidthCm', ylabel='Count'>
```



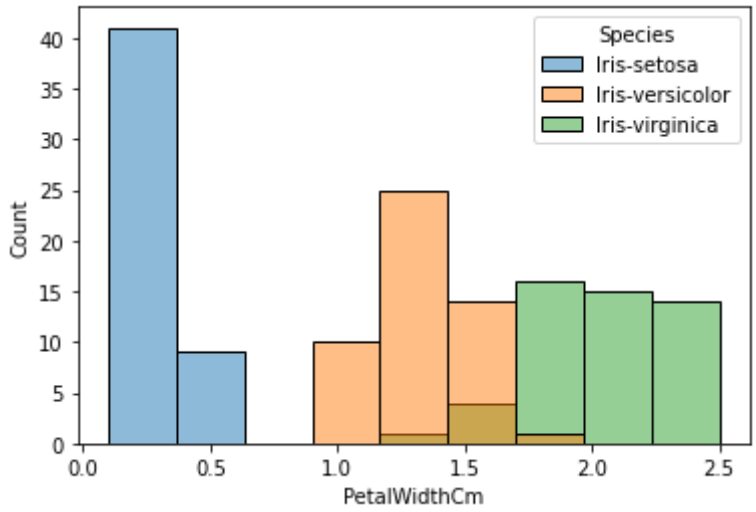
In [11]:
```
sns.histplot(data=iris, x="PetalLengthCm",hue="Species")
```

Out[11]:
```
<AxesSubplot:xlabel='PetalLengthCm', ylabel='Count'>
```



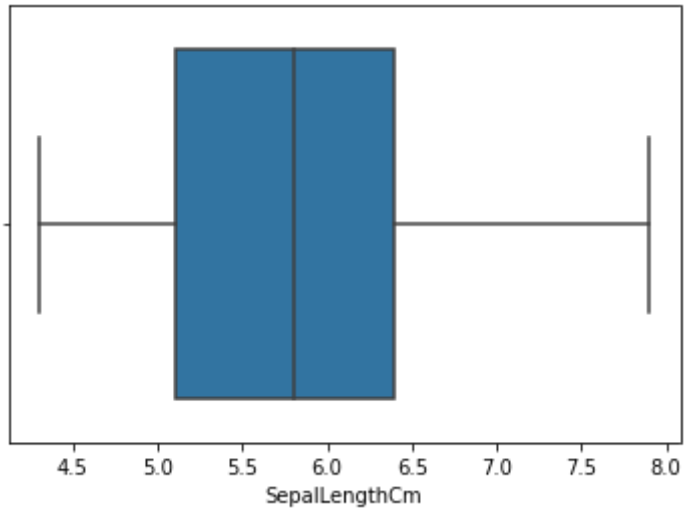
In [12]:
```
sns.histplot(data=iris, x="PetalWidthCm",hue="Species")
```

Out[12]:
```
<AxesSubplot:xlabel='PetalWidthCm', ylabel='Count'>
```

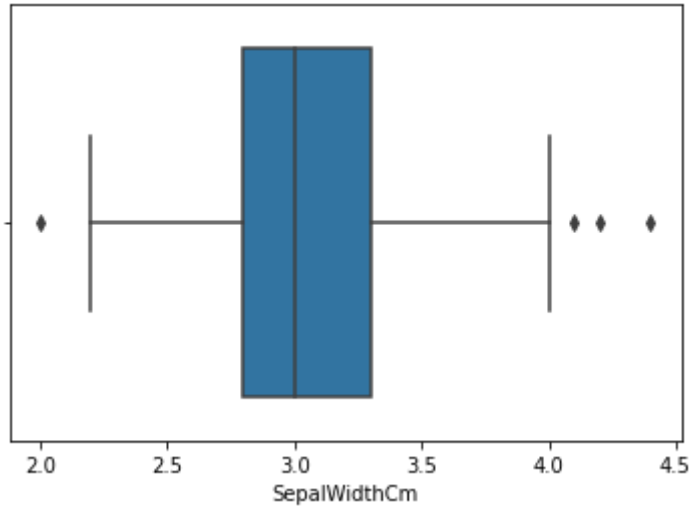


Create a box plot for each feature in the dataset.

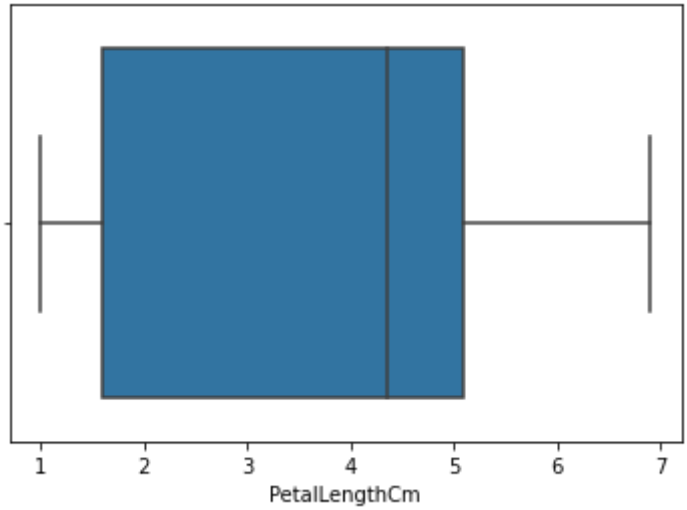
In [13]:
```
# sns.set_theme(style="darkgrid")

sns.boxplot(x=iris["SepalLengthCm"])
```



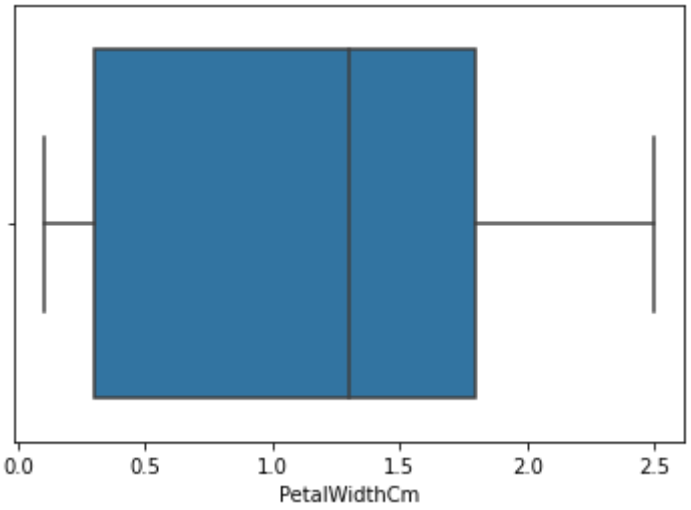
In [14]:
```
sns.boxplot(x=iris["SepalWidthCm"])
```



In [15]:
```
sns.boxplot(x=iris["PetalLengthCm"])
```



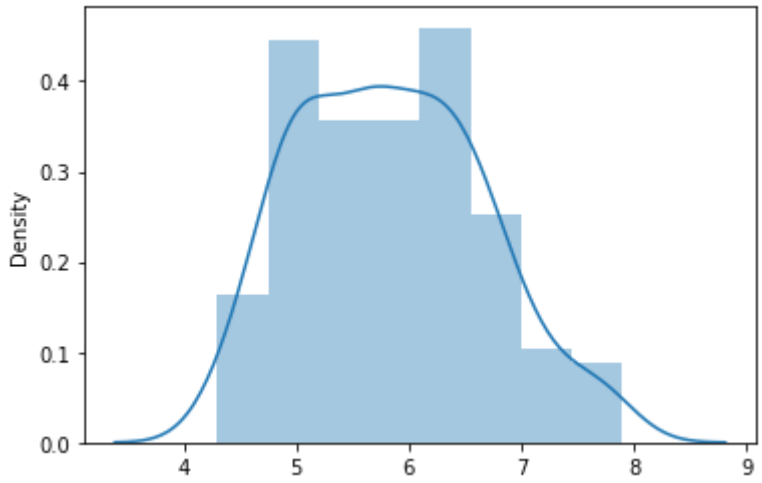
In [16]:
```
sns.boxplot(x=iris["PetalWidthCm"])
```



In [17]:
```
sns.distplot(x=iris.SepalLengthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

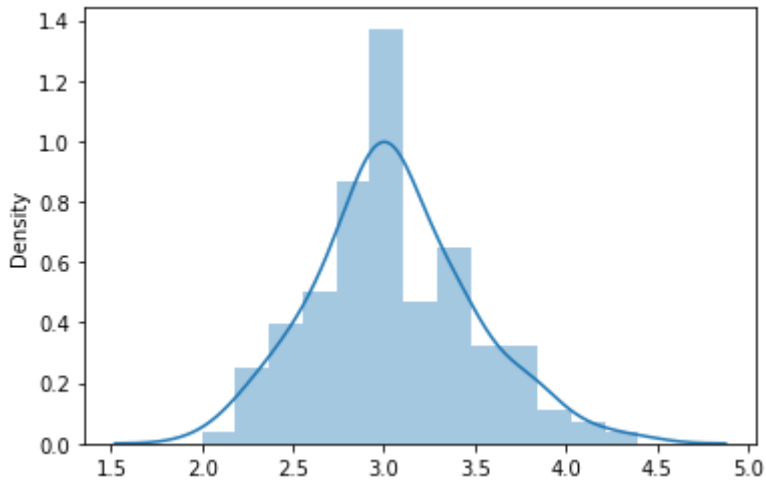
Out[17]:
```
<AxesSubplot:ylabel='Density'>
```



In [18]:
```
sns.distplot(x=iris.SepalWidthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

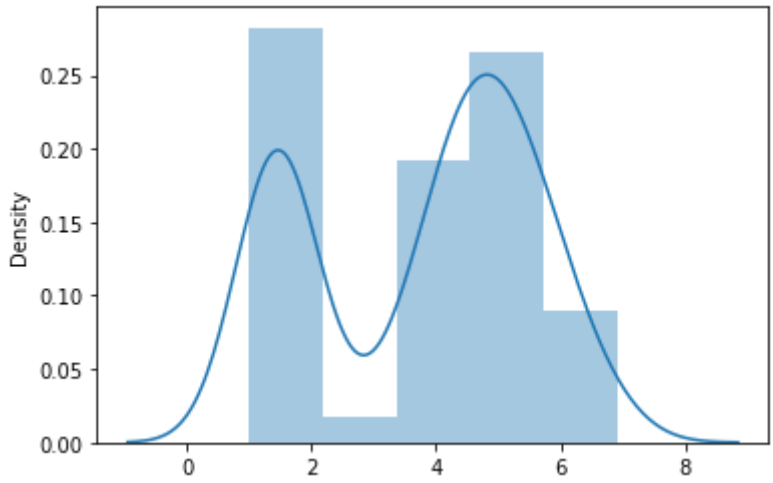
Out[18]:
```
<AxesSubplot:ylabel='Density'>
```



In [19]:
```
sns.distplot(x=iris.PetalLengthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

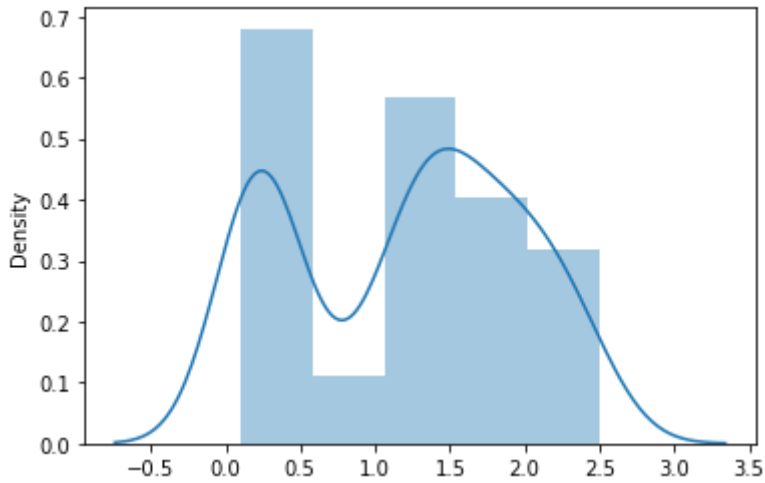
Out[19]:
```
<AxesSubplot:ylabel='Density'>
```



In [20]:
```
sns.distplot(x=iris.PetalWidthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[20]:
```
<AxesSubplot:ylabel='Density'>
```



In [21]:
```
Q1=iris['SepalWidthCm'].quantile(0.25)
Q3=iris['SepalWidthCm'].quantile(0.75)
IQR=Q3-Q1
print("IQR(", IQR, ") =", "Q3(", Q3, ")- Q1(", Q1, ")")
```

```
IQR( 0.5 ) = Q3( 3.3 )- Q1( 2.8 )
```

In [22]:
```
lower_limit=Q1-IQR
upper_limit=Q3+IQR
lower_limit,upper_limit
```

Out[22]:
```
(2.3, 3.8)
```

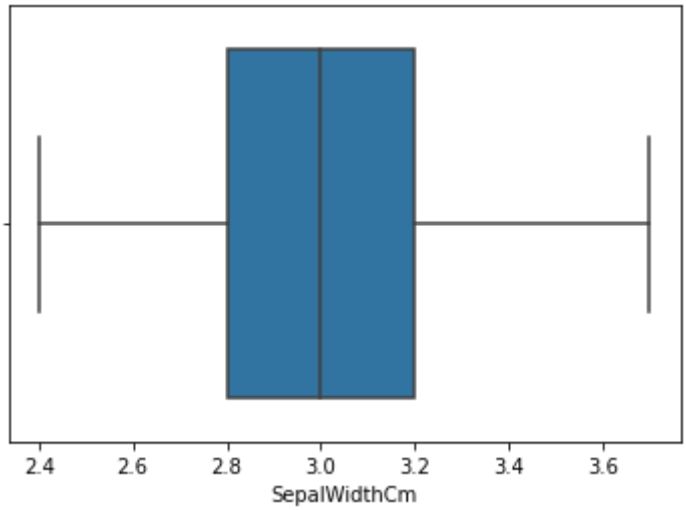
In [23]:
```
df_without_outliers=iris[(iris['SepalWidthCm']>lower_limit)&(iris['SepalWidthCm']<
df_without_outliers
```

Out[23]:

	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

130 rows × 6 columns

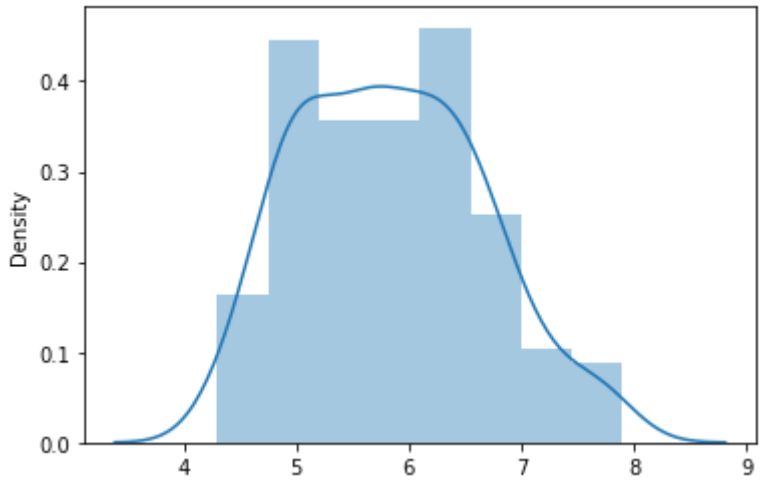
In [24]:
```
ax = sns.boxplot(x=df_without_outliers["SepalWidthCm"])
```



In [25]:
```
sns.distplot(x=iris.SepalLengthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

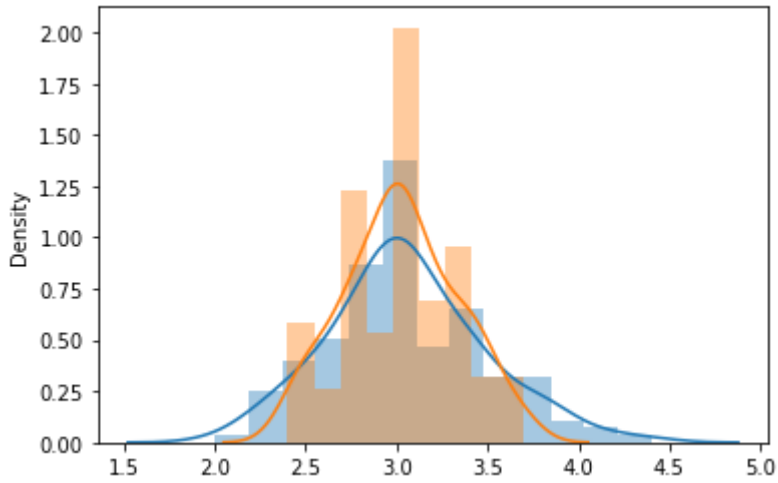
Out[25]:
```
<AxesSubplot:ylabel='Density'>
```



In [26]:
```
sns.distplot(x=iris.SepalWidthCm)

sns.distplot(x=df_without_outliers.SepalWidthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[26]:
```
<AxesSubplot:ylabel='Density'>
```

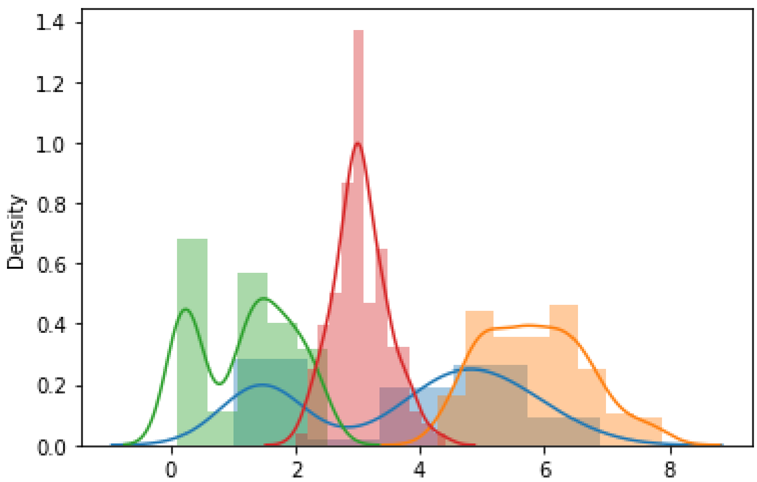


In [31]:
```
sns.distplot(x=iris.PetalLengthCm)
sns.distplot(x=iris.SepalLengthCm)
sns.distplot(x=iris.PetalWidthCm)
sns.distplot(x=iris.SepalWidthCm)
```

```
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/home/pict/.local/lib/python3.8/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[31]:

```
<AxesSubplot:ylabel='Density'>
```

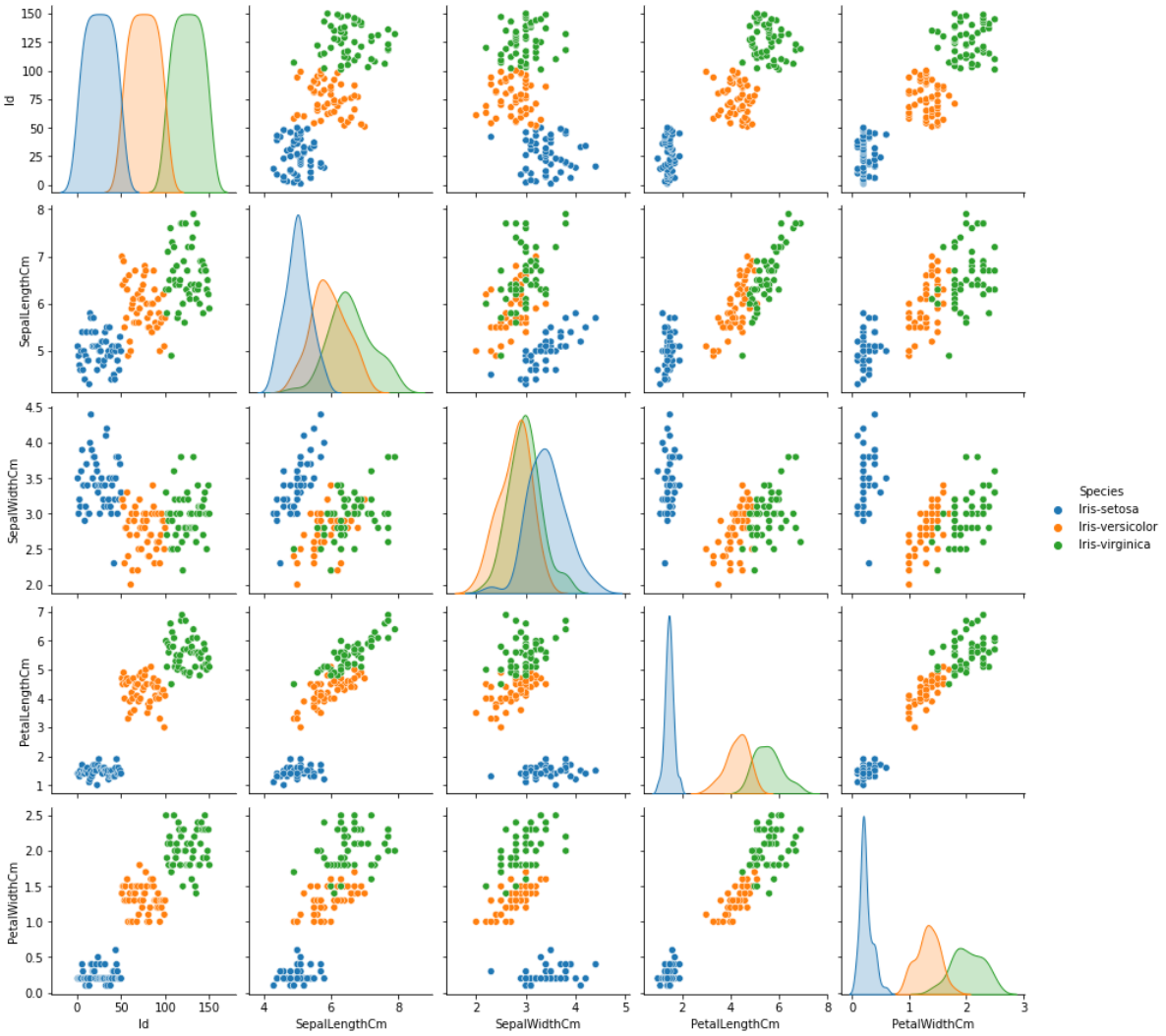


In [28]:

```python
sns.pairplot(data= iris, hue='Species')
```

Out[28]:

```
<seaborn.axisgrid.PairGrid at 0x7f754c40f4c0>
```


Id
SepalLengthCm
SepalWidthCm
PetalLengthCm
PetalWidthCm
Species
Iris-setosa
Iris-versicolor
Iris-virginica

In [ ]: