

In [4]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [8]:

```
iris_df = pd.read_csv("iris.csv")
print("the data has been successfully load.")
```

the data has been successfully load.

In [9]:

```
iris_df
```

Out[9]:

	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]:

```
iris_df.shape
```

Out[10]:

(150, 6)

In [11]:

```
#check basic information of data
iris_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 [12]:

```
iris_df.describe()
```

Out[12]:

	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 [13]:

```
iris_df.isnull().sum()
```

Out[13]:

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

In [14]:

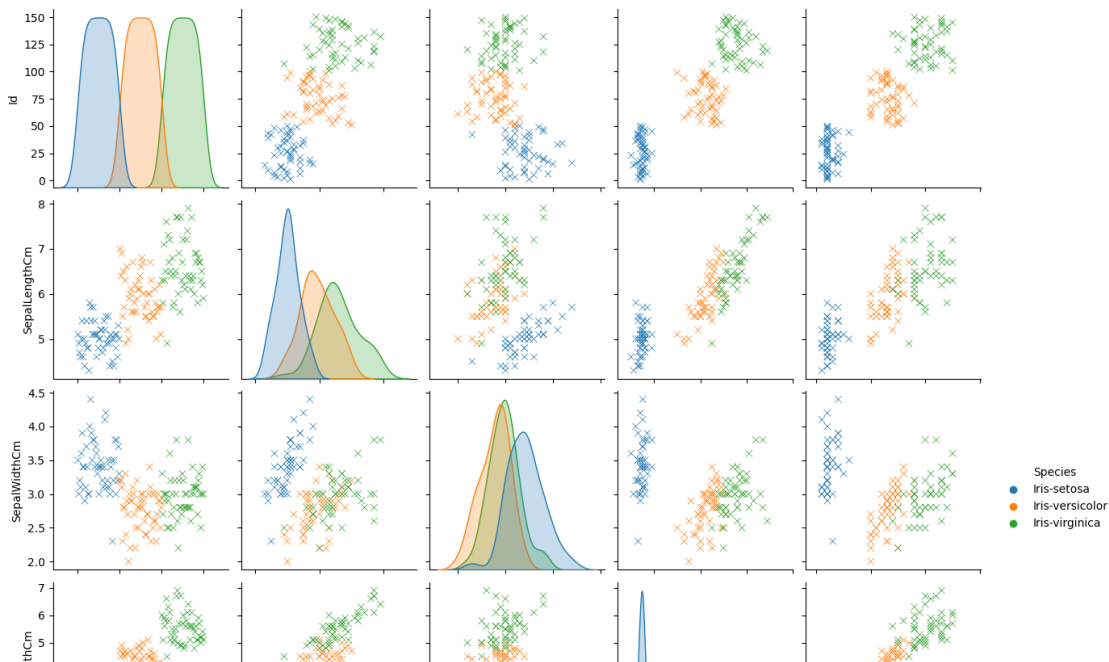
```
print("unique number of values in dataset Species:",iris_df["Species"].nunique())
print("Unique number Species in iris datasets:",iris_df["Species"].unique())
```

unique number of values in dataset Species: 3

Unique number Species in iris datasets: ['Iris-setosa' 'Iris-versicolor'
'Iris-virginica']

In [15]:

```
sns.pairplot(iris_df,hue="Species",markers="x")
plt.show()
```



In [18]:

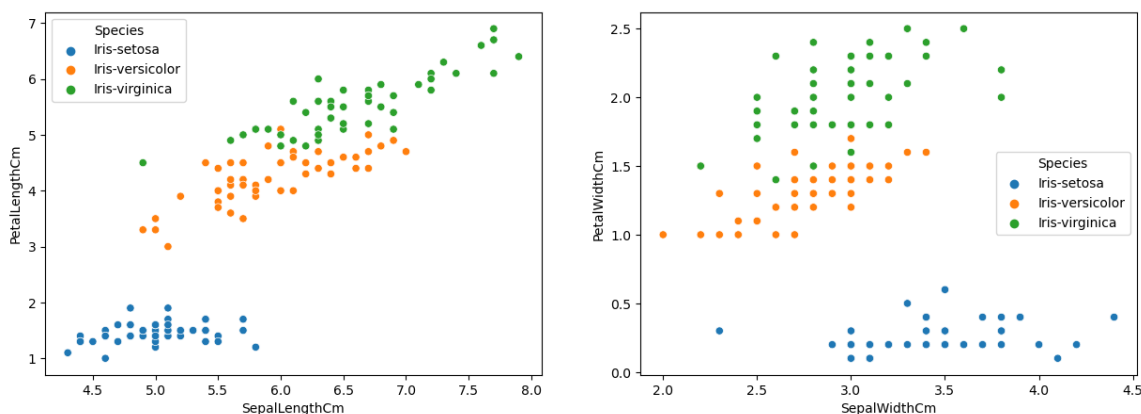
```
plt.figure(figsize=(15,5))
plt.subplot(1,2,1)

sns.scatterplot (x='SepalLengthCm',y='PetalLengthCm',data=iris_df, hue='Species')

plt.subplot(1,2,2)

sns.scatterplot(x='SepalWidthCm',y='PetalWidthCm',data=iris_df, hue='Species')

plt.show()
```



In [19]:

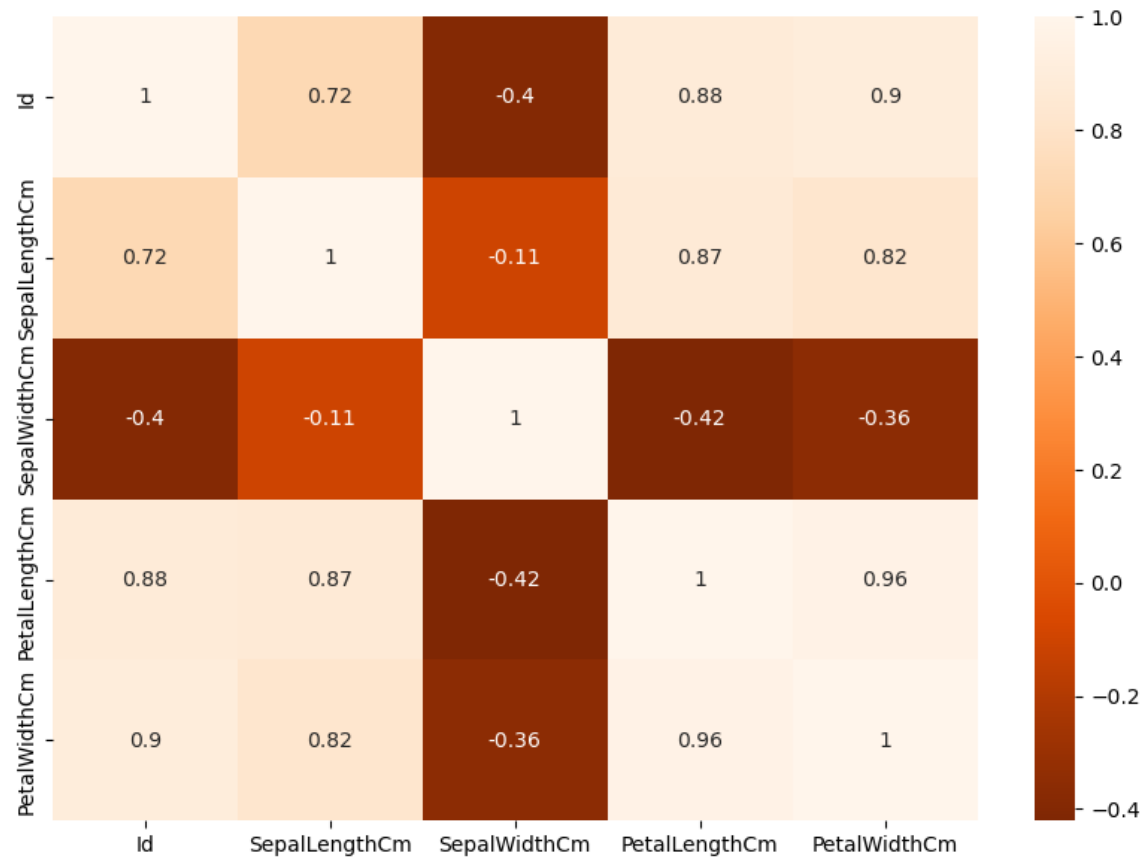
```
iris_df.corr()
```

Out[19]:

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm
Id	1.000000	0.716676	-0.397729	0.882747	0.899759
SepalLengthCm	0.716676	1.000000	-0.109369	0.871754	0.817954
SepalWidthCm	-0.397729	-0.109369	1.000000	-0.420516	-0.356544
PetalLengthCm	0.882747	0.871754	-0.420516	1.000000	0.962757
PetalWidthCm	0.899759	0.817954	-0.356544	0.962757	1.000000

In [20]:

```
plt.figure(figsize=(10,7))
sns.heatmap(iris_df.corr(),annot=True,cmap="Oranges_r")
plt.show()
```

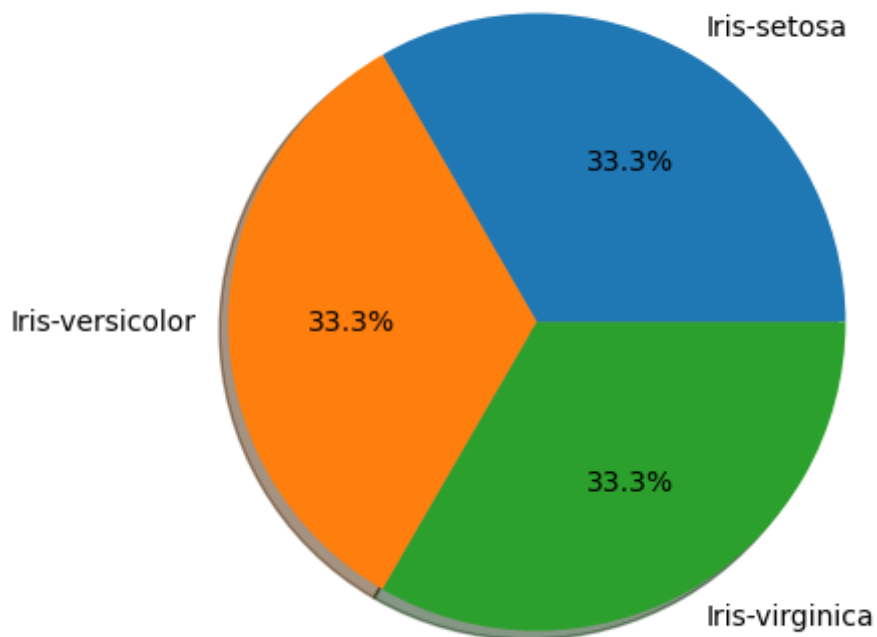


In [24]:

```
# Check value counts
```

```
iris_df["Species"].value_counts().plot(kind="pie", autopct = "%1.1f%%", shadow=True, figsize=(10, 10))  
plt.title("Percentage values in each Species", fontsize = 12, c = "g")  
plt.ylabel("", fontsize=10, c="r")  
plt.show()
```

Percentage values in each Species



In [26]:

```
sns.scatterplot(iris_df["SepalLengthCm"], iris_df["SepalWidthCm"], hue=iris_df["Species"])  
plt.show()
```

