

```
In [1]: # Exploratory Data Analysis (EDA) on Iris Dataset
```

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("Iris.csv")
```

```
In [80]: #Top 5 values in the data
df.head()
```

```
Out[80]:
```

	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

```
In [78]: #Bottom 5 values in the data
df.tail()
```

```
Out[78]:
```

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
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

```
In [26]: #Total number of Columns and Rows  
df.shape
```

```
Out[26]: (150, 6)
```

```
In [25]: #Data types of the columns  
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 [77]: #To check null values in the data  
df.isnull().sum()
```

```
Out[77]: Id                0  
SepalLengthCm            0  
SepalWidthCm             0  
PetalLengthCm            0  
PetalWidthCm             0  
Species                  0  
dtype: int64
```

```
In [76]: #To check Duplicates  
df.duplicated().value_counts()
```

```
Out[76]: False      150  
dtype: int64
```

```
In [75]: #Total count for each species
df["Species"].value_counts()
```

```
Out[75]: Iris-setosa      50
Iris-versicolor    50
Iris-virginica     50
Name: Species, dtype: int64
```

```
In [73]: #Statistical Analysis
df.describe()
```

```
Out[73]:
```

	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 [72]: #Unique values in column
df["Species"].unique()
```

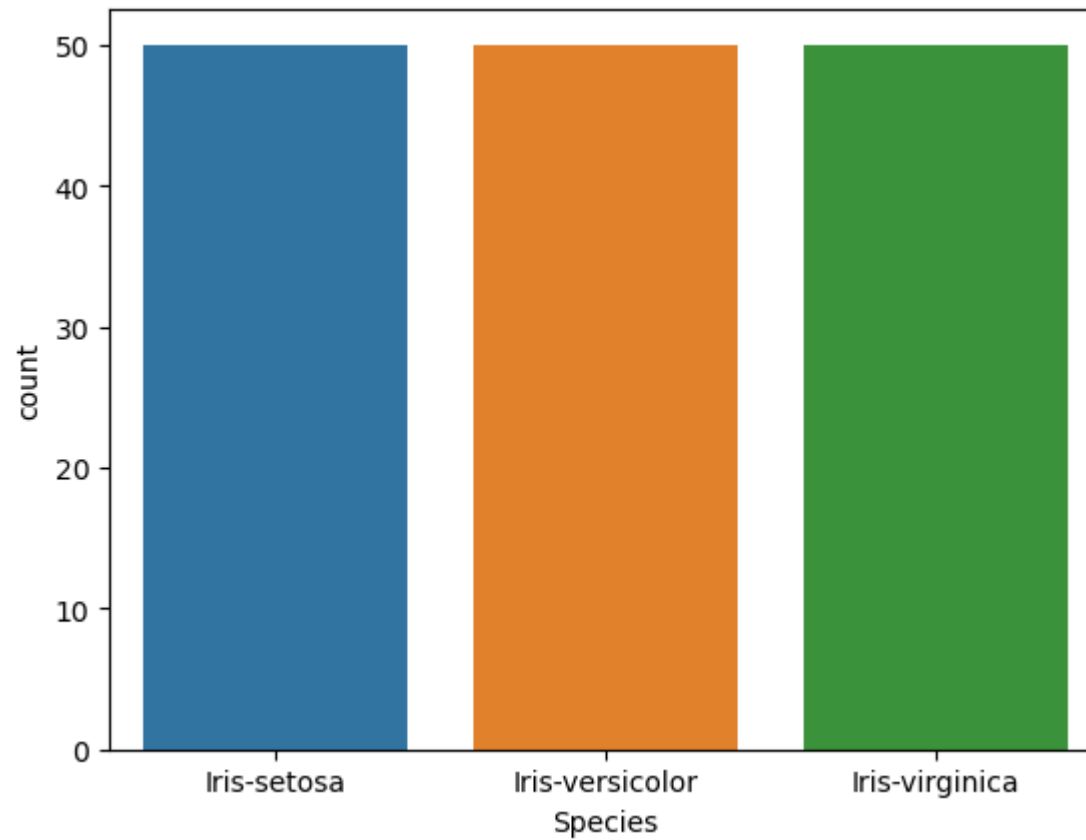
```
Out[72]: array(['Iris-setosa', 'Iris-versicolor', 'Iris-virginica'], dtype=object)
```

```
In [71]: #No of Unique values in column
df["Species"].nunique()
```

```
Out[71]: 3
```

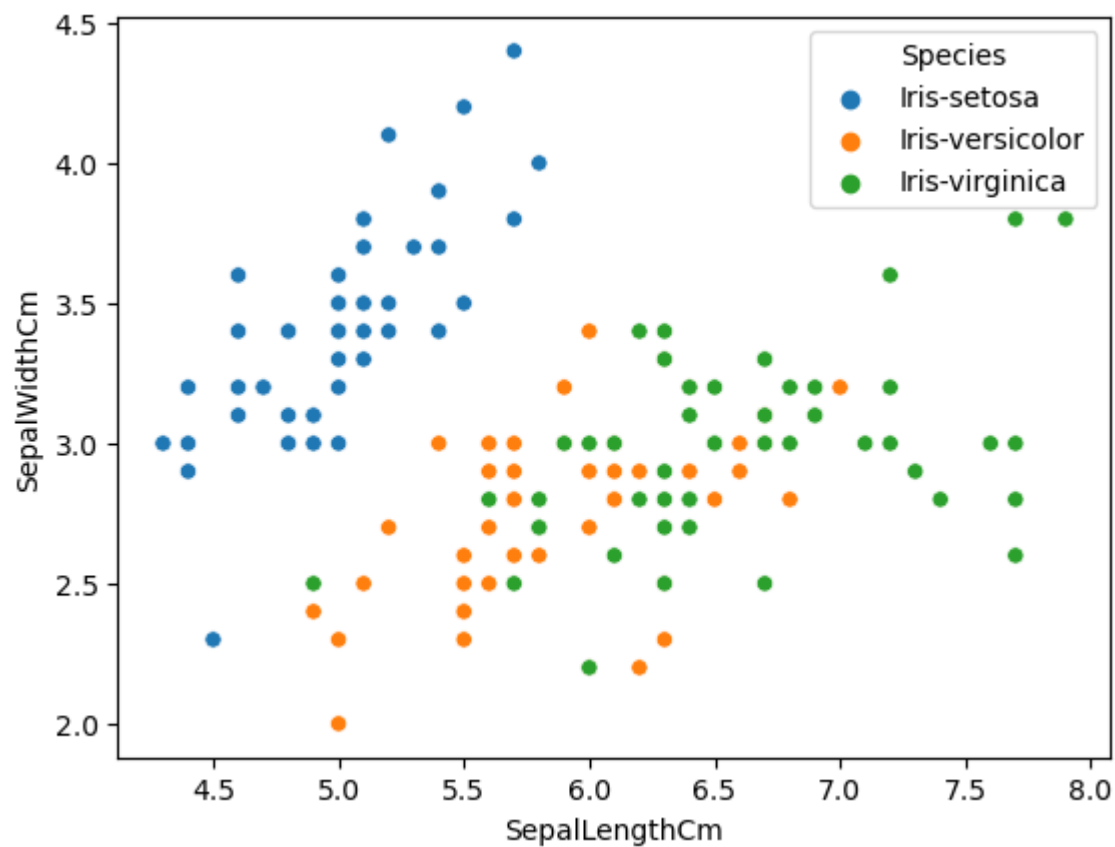
```
In [70]: #Visualization  
sns.countplot(x="Species",data=df)
```

```
Out[70]: <Axes: xlabel='Species', ylabel='count'>
```



```
In [69]: #Comparing SepalLength and SepalWidth  
sns.scatterplot(x="SepalLengthCm",y="SepalWidthCm",hue="Species",data=df )
```

```
Out[69]: <Axes: xlabel='SepalLengthCm', ylabel='SepalWidthCm'>
```



```
In [ ]:
```

```
In [ ]:
```

