

# FINDING INSIGHTS ON DATA USING PAIRPLOT

## IMPORTING LIBRARIES

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

## LOADING THE DATA INTO A DATAFRAME

```
In [2]: data = pd.read_csv('iris_dataset.csv')
```

## GETTING TO KNOW ABOUT OUR DATA

```
In [3]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 105 entries, 0 to 104
Data columns (total 5 columns):
 #   Column          Non-Null Count  Dtype  
---  --
 0   sepal_length    105 non-null   float64
 1   sepal_width     101 non-null   float64
 2   petal_length    97 non-null    float64
 3   petal_width     105 non-null   float64
 4   species         105 non-null   object  
dtypes: float64(4), object(1)
memory usage: 4.2+ KB
```

```
In [4]: data.shape
```

```
Out[4]: (105, 5)
```

```
In [5]: data.head()
```

```
Out[5]:
```

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.045070	2.508203	3.018024	1.164924	Iris-versicolor
1	6.325517	2.115481	4.542052	1.413651	Iris-versicolor
2	5.257497	3.814303	1.470660	0.395348	Iris-setosa
3	6.675168	3.201700	5.785461	2.362764	Iris-virginica
4	5.595237	2.678166	4.077750	1.369266	Iris-versicolor

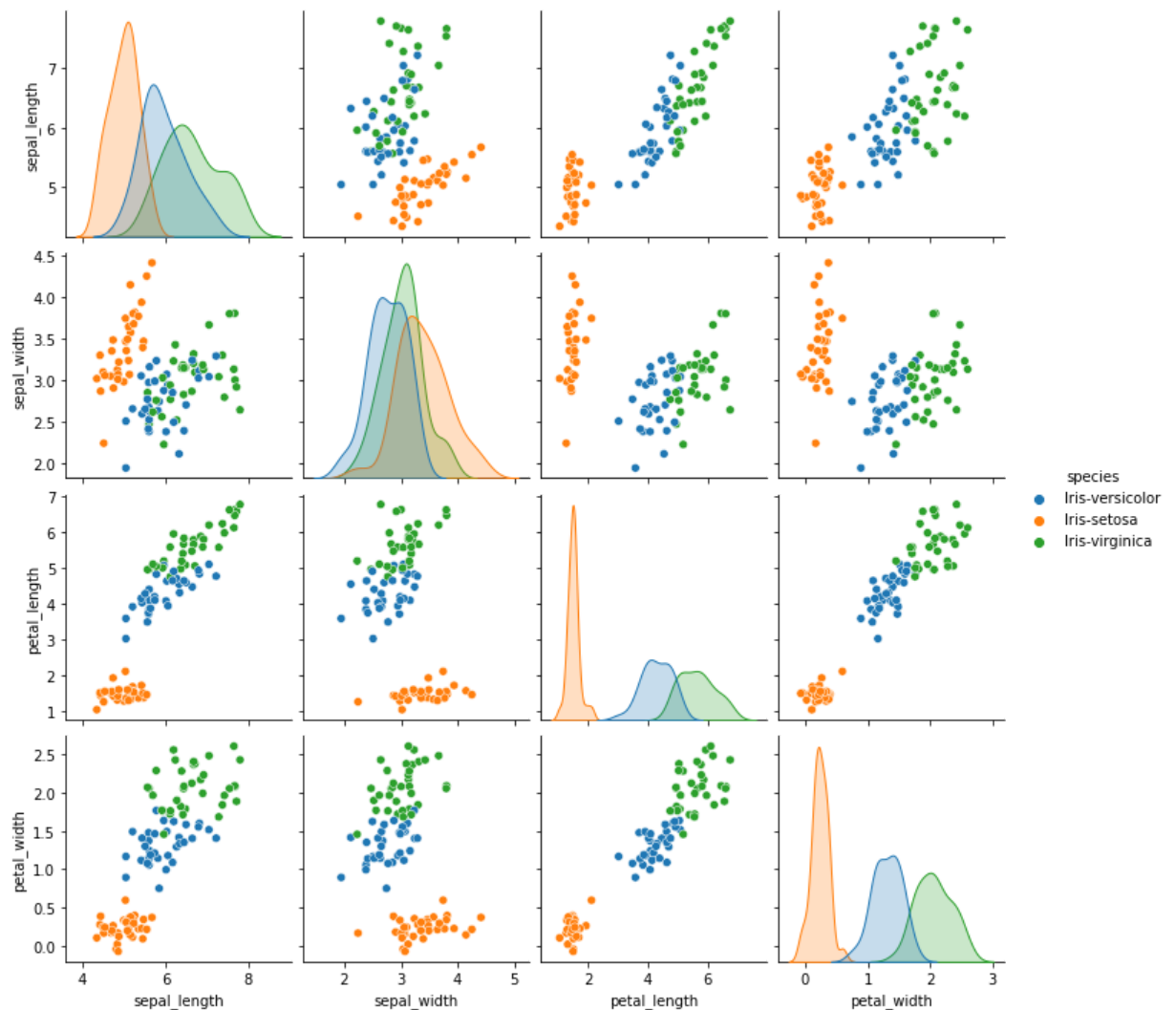
```
In [6]: data.describe()
```

```
Out[6]:
```

	sepal_length	sepal_width	petal_length	petal_width
<b>count</b>	105.000000	101.000000	97.000000	105.000000
<b>mean</b>	5.858909	3.059083	3.812370	1.199708
<b>std</b>	0.861638	0.455116	1.793489	0.787193
<b>min</b>	4.344007	1.946010	1.033031	-0.072203
<b>25%</b>	5.159145	2.768688	1.545136	0.333494
<b>50%</b>	5.736104	3.049459	4.276817	1.331797
<b>75%</b>	6.435413	3.290318	5.094427	1.817211
<b>max</b>	7.795561	4.409565	6.768611	2.603123

***PAIR PLOT : VISUALIZING THE DISTRIBUTION OF A VARIABLE BETWEEN MULTIPLE VARIABLES SEPARATELY***

```
In [7]: pplot=sns.pairplot(data,hue='species')
```



We plotted the distribuion of data of all the features in a pair plot and made use of the parameter called **"Hue"** to distinguish the data with the help of the dependent variable **"species"**.

Here, we have considered all the features from the above pair plot as there is a significant amount of relationship among them, which can be used in increasing the accuracy of the model.