56. Agglomerate Hierarchical (Practical)

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

In [3]:
```
dataset = pd.read_csv(r'Data/iris_raw.csv')
dataset.head(3)
```

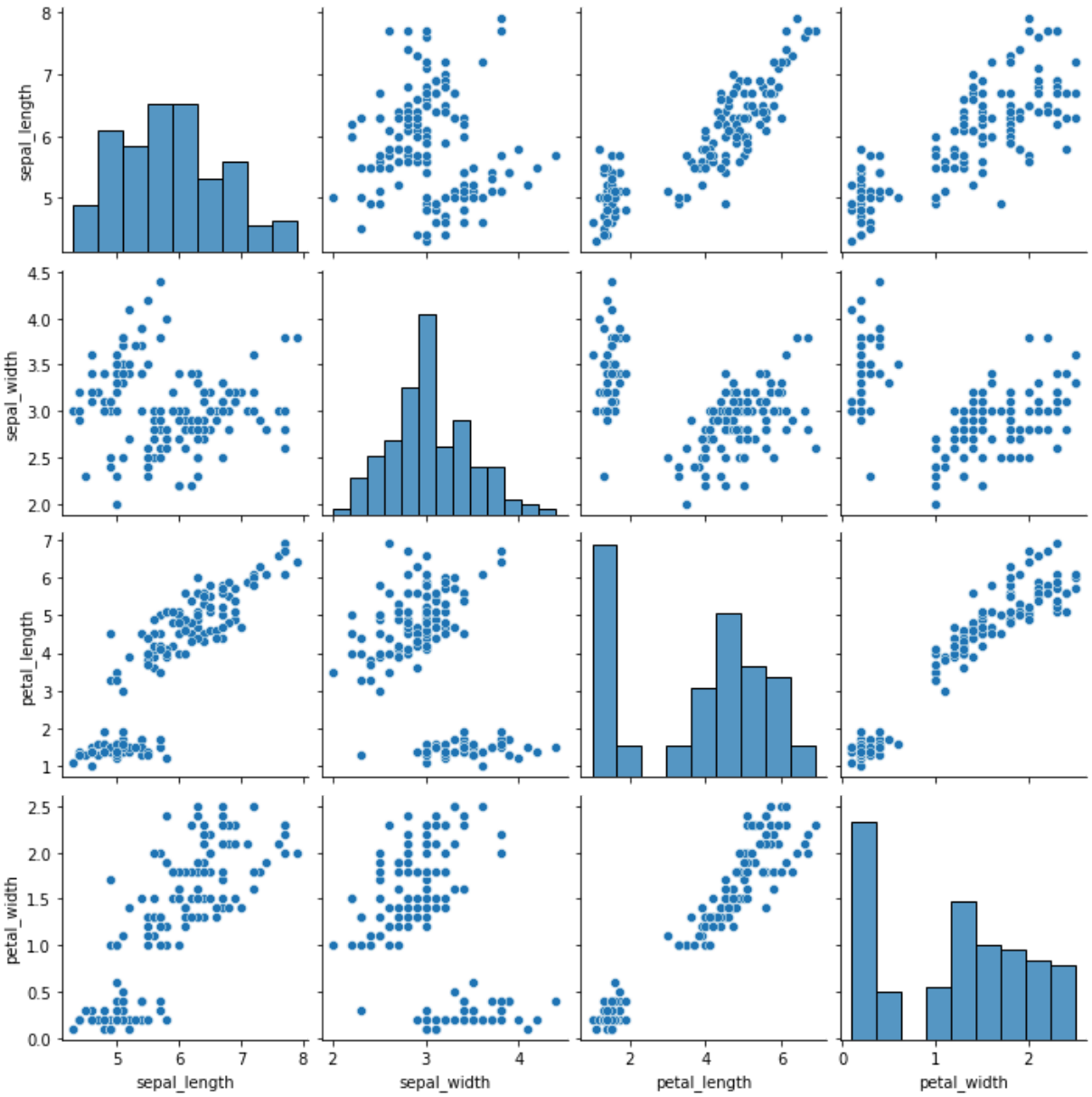
Out[3]:

	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2

As agglomerate clustering works on **linearly separable data**, so we will see if our data is linear or not through graph

In [4]:
```
sns.pairplot(data=dataset)
plt.show()
```

```
C:\Users\rashi\AppData\Local\Programs\Python\Python39\lib\site-packages\seaborn\axisgrid.py:123: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



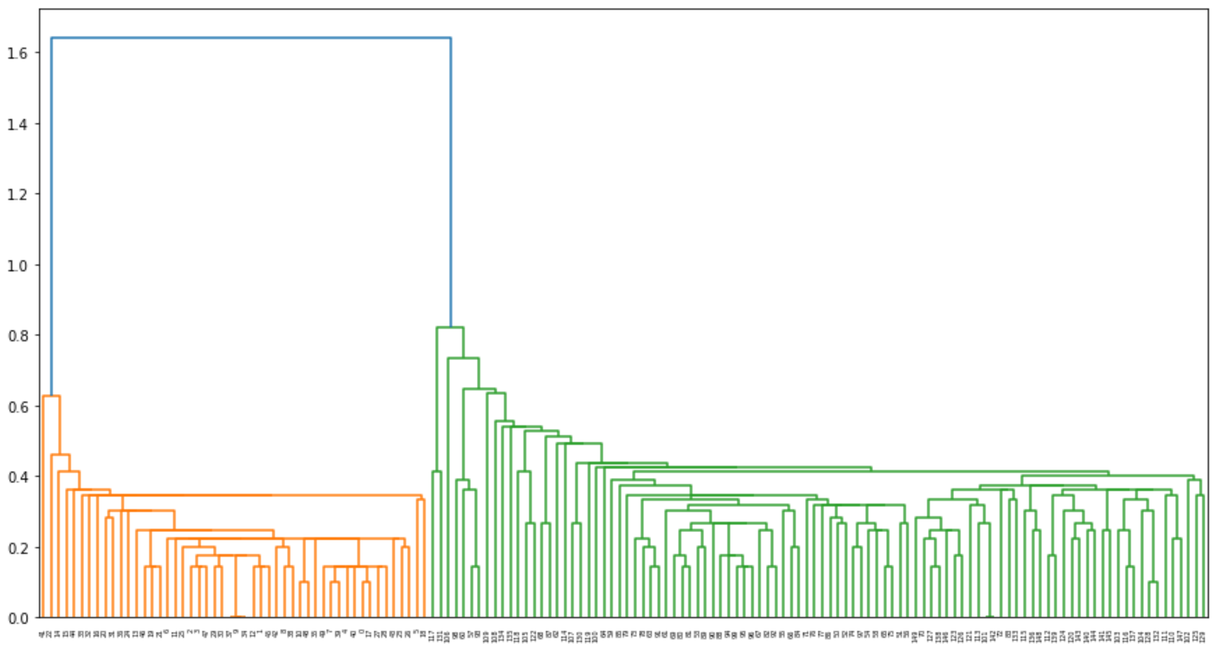
Make Dendrogram

SciPy Library is needed for making dendrogram

In [6]:
```
import scipy.cluster.hierarchy as sc
```

We will need **Linkage** fro making dendrogram

In [10]:
```
'''Z : ndarray
    The linkage matrix encoding the hierarchical clustering to
    render as a dendrogram. See the ``linkage`` function for more
    information on the format of ``Z``.'''
plt.figure(figsize=(15,8))
sc.dendrogram(sc.linkage(dataset, method='single', metric='euclidean'))
plt.savefig(r'Generated_images/dendrogram.jpg')
plt.show()
```



Dendrogram is showing two clusters only in the data

In [11]:
```python
from sklearn.cluster import AgglomerativeClustering
```

In [13]:
```python
ac = AgglomerativeClustering(n_clusters=2, linkage='single')
ac.fit_predict(dataset)
```

Out[13]:
```
array([1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], dtype=int64)
```

In [14]:
```python
dataset['Predict'] = ac.fit_predict(dataset)
```

In [15]:
```python
dataset
```

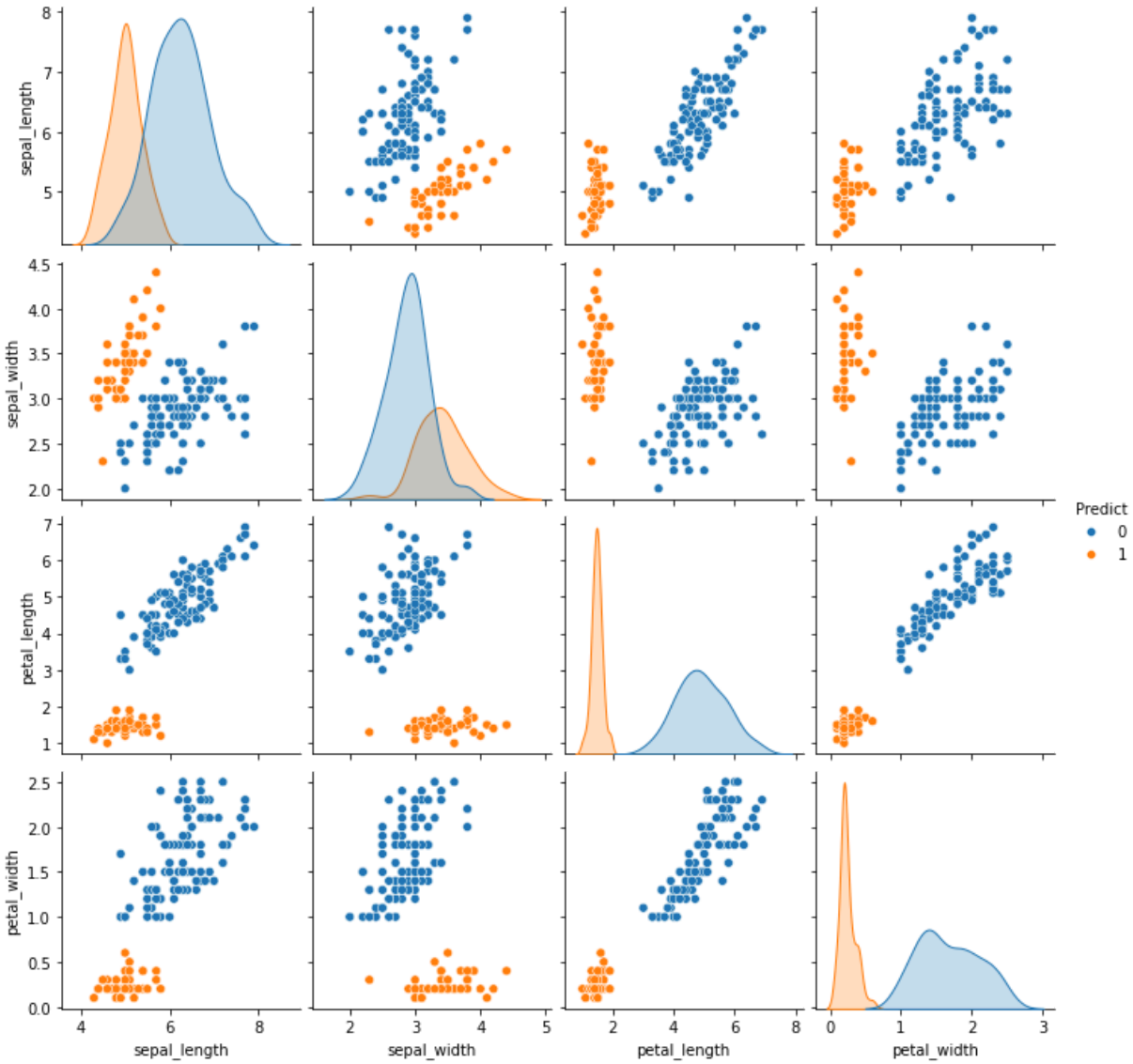
Out[15]:

	sepal_length	sepal_width	petal_length	petal_width	Predict
0	5.1	3.5	1.4	0.2	1
1	4.9	3.0	1.4	0.2	1
2	4.7	3.2	1.3	0.2	1
3	4.6	3.1	1.5	0.2	1
4	5.0	3.6	1.4	0.2	1
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	0
146	6.3	2.5	5.0	1.9	0
147	6.5	3.0	5.2	2.0	0
148	6.2	3.4	5.4	2.3	0
149	5.9	3.0	5.1	1.8	0

150 rows × 5 columns

In [16]:
```
sns.pairplot(data=dataset, hue='Predict')
plt.show()
```

```
C:\Users\rashi\AppData\Local\Programs\Python\Python39\lib\site-packages\seaborn\axisgrid.py:123: UserWarning: The figure layout has changed to tight
  self._figure.tight_layout(*args, **kwargs)
```



Prediction is also showing two clusters only in the data

In [ ]: