In [51]:
```python
#Step1: Import all the important Libraries

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [8]:
```python
#Step2: Load the CSV file in the irisDataset dataframe


irisDataset = pd.read_csv(r"C:\Users\ayesh\OneDrive\Desktop\BI Program\1 SEM\CST2101-Business Intelligence Prog
```

In [32]:
```python
#Step3: Let's get a quick idea about what the data looks like
irisDataset.head()
```

Out[32]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

In [36]:
```python
#Step4: Let's understand what are the datatypes
irisDataset.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

In [50]:
```python
#Step5: Let's find out if there are any missing values
irisDataset.isnull().sum()
```

Out[50]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

In [39]:
```python
#Step6: Use shape attribute to know the dimensions( rows, columns)

print(irisDataset.shape)
```

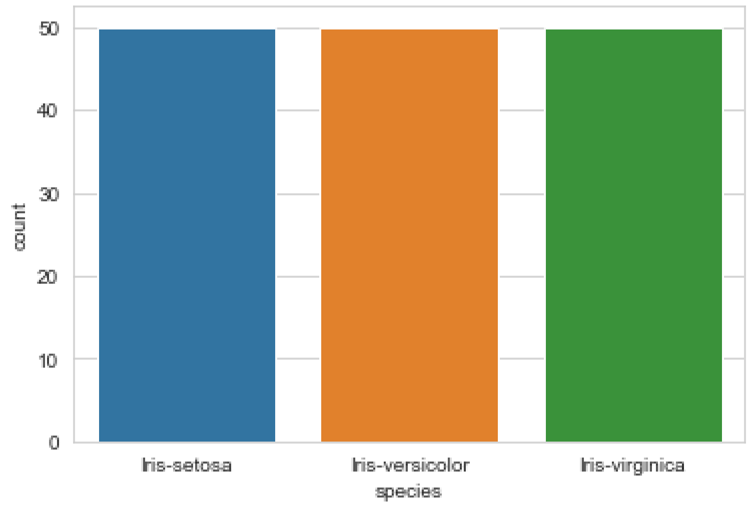
```
(150, 5)
```

In [40]:
```python
#Step7 : Find out what columns/features are present in the dataset and also plot them

print(irisDataset.columns)
sns.countplot(x='species',data=irisDataset)
```

```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

Out[40]:
```
<AxesSubplot:xlabel='species', ylabel='count'>
```



In [41]:
```python
#Step8: Use describe to find out the summary statistics of the data
print(irisDataset.describe())
```

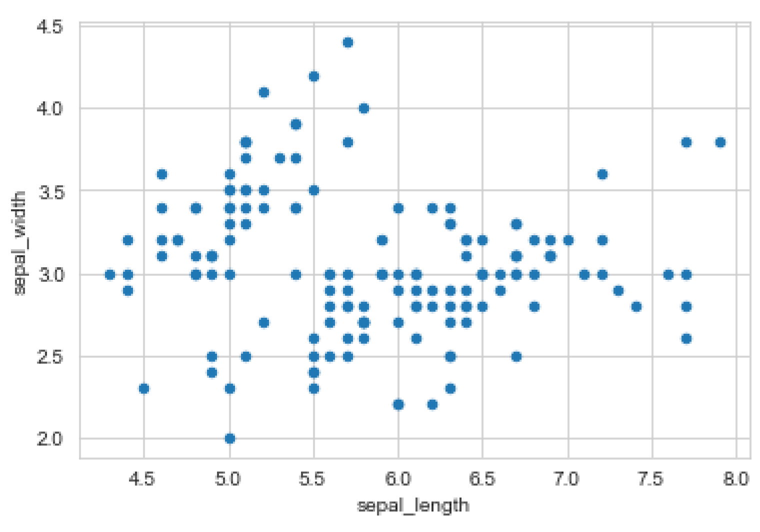
```
       sepal_length  sepal_width  petal_length  petal_width
count    150.000000   150.000000    150.000000   150.000000
mean       5.843333     3.054000      3.758667     1.198667
std        0.828066     0.433594      1.764420     0.763161
min        4.300000     2.000000      1.000000     0.100000
25%        5.100000     2.800000      1.600000     0.300000
50%        5.800000     3.000000      4.350000     1.300000
75%        6.400000     3.300000      5.100000     1.800000
max        7.900000     4.400000      6.900000     2.500000
```

In [42]:
```python
#Step9: Let's try to understand what are the different class labels that are present in rhe dataset.

print(irisDataset["species"].value_counts())
```

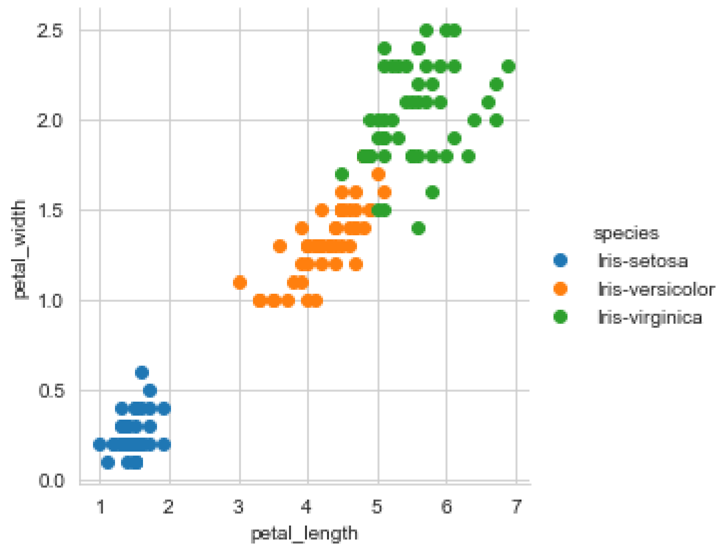
```
Iris-virginica     50
Iris-setosa        50
Iris-versicolor    50
Name: species, dtype: int64
```

In [43]:
```python
#Step10: Visulaizing this dataset with the help of a scatter plot by taking'sepal_length' as x-axis and 'sepal_
irisDataset.plot(kind = "scatter", x = "sepal_length", y = "sepal_width")
plt.show()
```

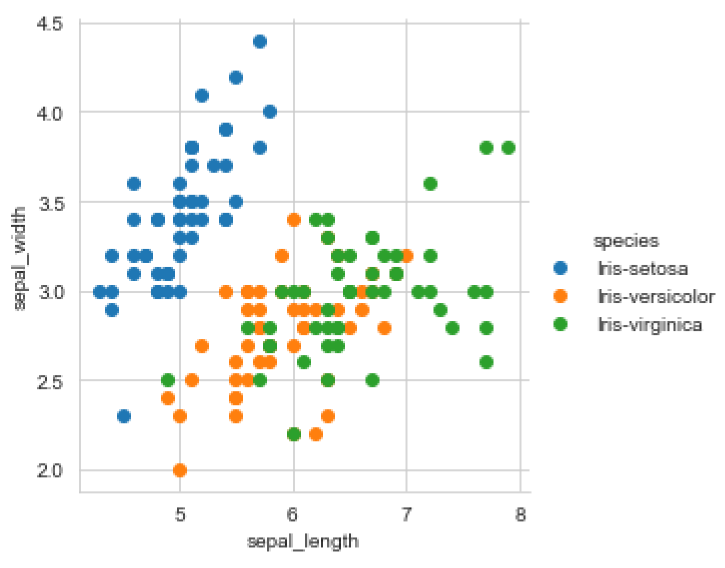


In [47]:
```python
#Step11:Since the previous plot was not very helpful ,Lets visulaize the dataset with the help of seaborn libra

sns.set_style("whitegrid")
sns.FacetGrid(irisDataset, hue = "species", height = 4) \
   .map(plt.scatter, "petal_length", "petal_width") \
   .add_legend()
plt.show()
```



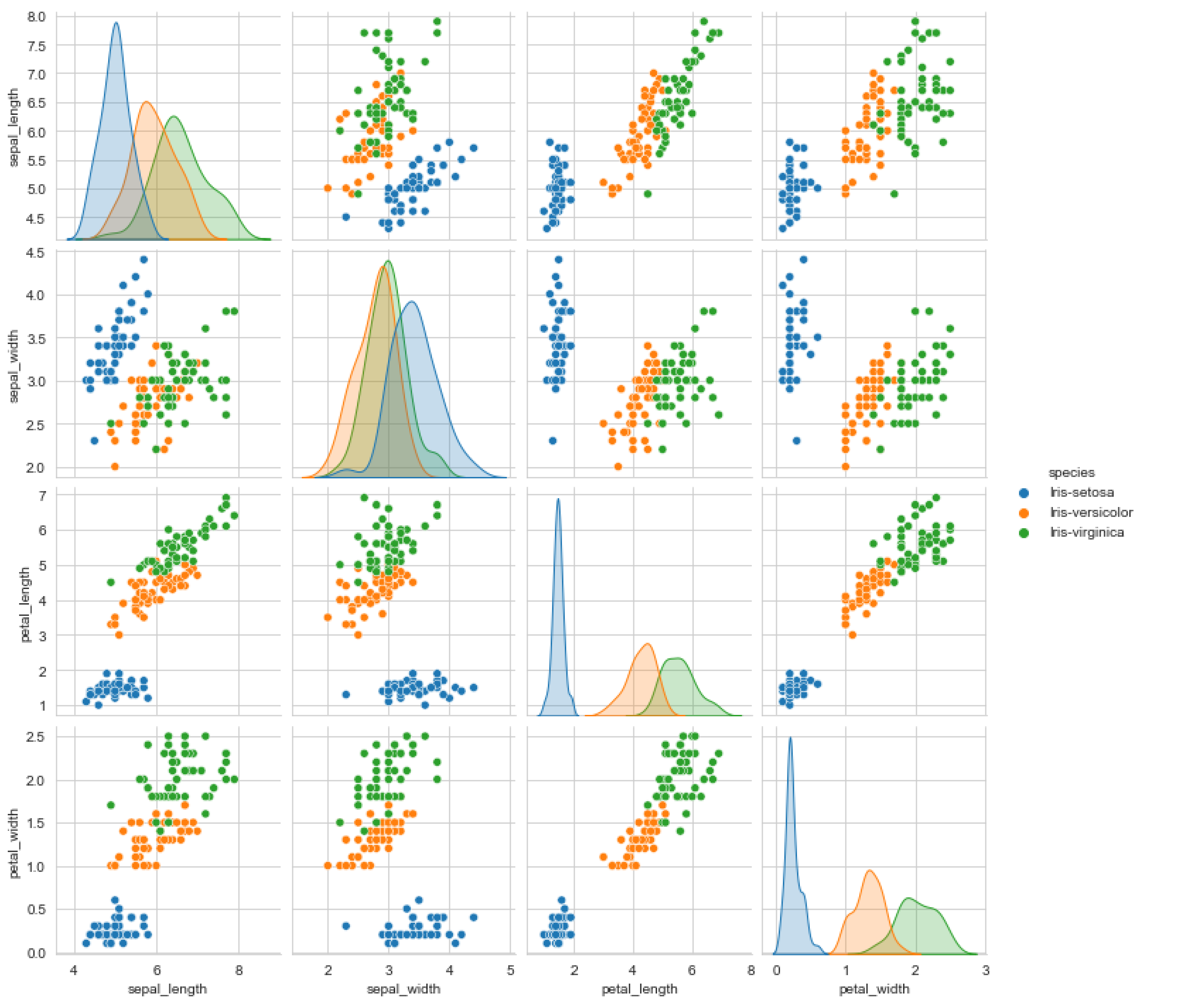
In [46]:
```python
#Step12: Let's try with 'sepal_length', 'sepal_width'
sns.set_style("whitegrid")
sns.FacetGrid(irisDataset, hue = "species", height = 4) \
   .map(plt.scatter, 'sepal_length', 'sepal_width') \
   .add_legend()
plt.show()
```



In [49]:
```python
#Step13: Let's try to visulaize the different combinations between the remaining features by using pairplots

sns.pairplot(irisDataset, hue='species')
```

Out[49]:
```
<seaborn.axisgrid.PairGrid at 0x1c5db4cadf0>
```



Conclusion:

- The seaborn library helps us visualize the data points in a more appealing manner comapred to a scatter plot.
- By visulaizing a 2D plot between Sepal length Vs Sepal Width and Petal Length Vs Petal Width we see that the datapoints - separated well when we consider the Petal Length Vs Petal Width
- When we try to extend our analysis using pair plots which is used to gain insights into the relationships between the four features, it becomes evident that Petal Length and Petal Width continue to stand out as superior features for classifying the data.

In [ ]: