Program 2 - Write a program of KMeans clustring using ml techinque

In [76]:

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

In [38]:

```
df=sns.load_dataset('iris')
```

In [39]:

```
df.head()
```

Out[39]:

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

In [40]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
sepal_length    150 non-null float64
sepal_width     150 non-null float64
petal_length    150 non-null float64
petal_width     150 non-null float64
species         150 non-null object
dtypes: float64(4), object(1)
memory usage: 5.9+ KB
```

In [41]:

```
df.describe()
```

Out[41]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
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 [45]:

```
df.isnull().sum()
```

Out[45]:

```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

In [48]:

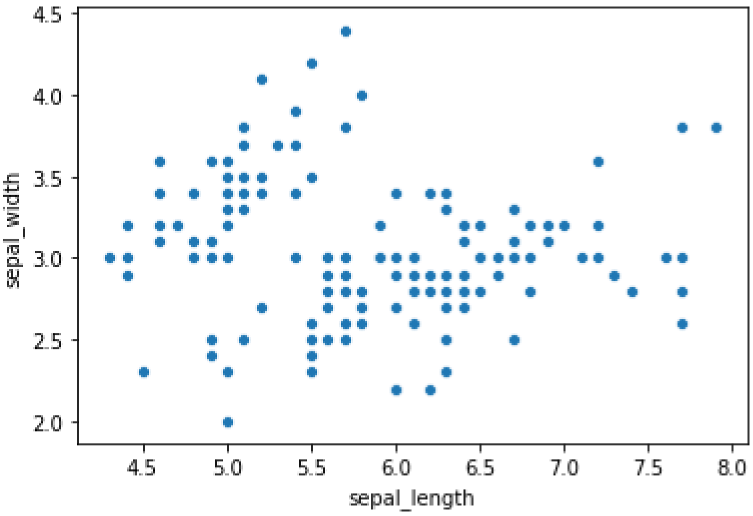
```
df.duplicated().sum()
```

Out[48]:

```
1
```

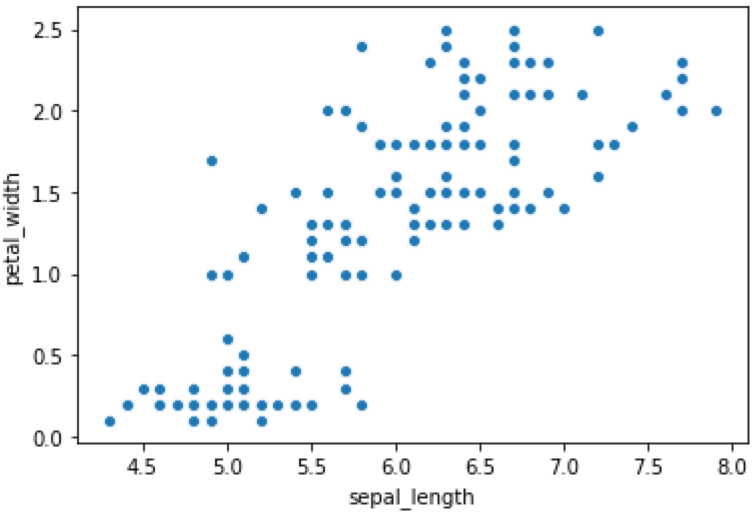
In [74]:

```
sns.scatterplot(x='sepal_length',y='sepal_width',data=d)
plt.show()
```

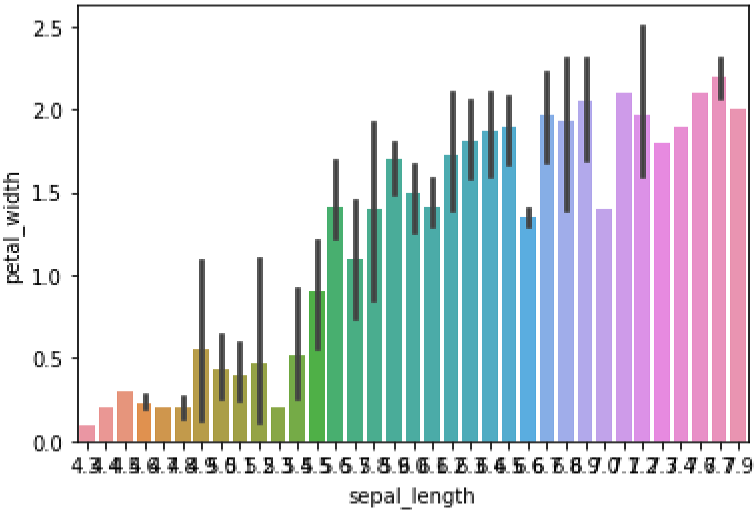


In [63]:

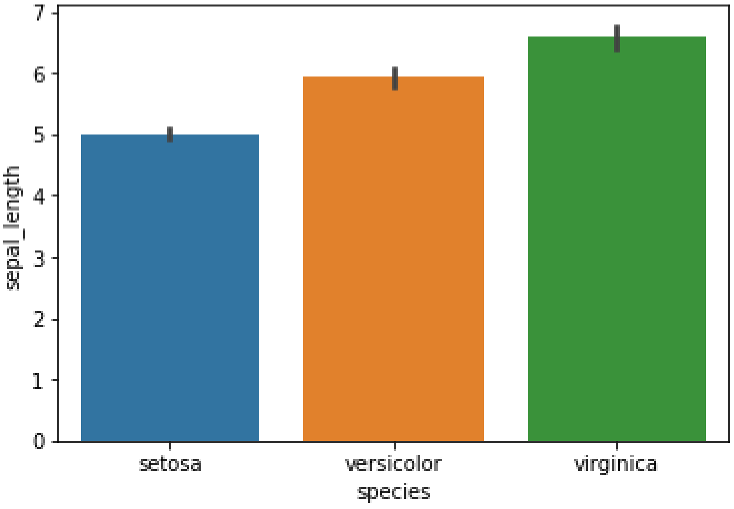
```
sns.scatterplot(x='sepal_length',y='petal_width',data=df)
plt.show()
```



In [64]:
```
sns.barplot(x='sepal_length',y='petal_width',data=df)
plt.show()
```



In [66]:
```
sns.barplot(y='sepal_length',x='species',data=df)
plt.show()
```



In [19]:
```
df = df.drop(columns=['species'])
```

In [27]:
```
from sklearn.cluster import KMeans
```

In [28]:
```
kmean=KMeans(n_clusters=3,max_iter=150,random_state=20)
```

In [30]:
```
kmean
```

Out[30]:
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=150,
       n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=20, tol=0.0001, verbose=0)
```

In [31]:

```
kmean.fit(df)
```

Out[31]:

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=150,
       n_clusters=3, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=20, tol=0.0001, verbose=0)
```

In [32]:

```
 kmean.labels_
```

Out[32]:

```
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, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 0, 0, 0, 0, 2, 0, 0, 0,
       0, 0, 0, 2, 2, 0, 0, 0, 0, 2, 0, 2, 0, 2, 0, 0, 2, 2, 0, 0, 0, 0,
       0, 2, 0, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 0, 2, 0, 0, 2])
```

In [78]:

```
df['cluster']= kmean.labels_
```

In [81]:

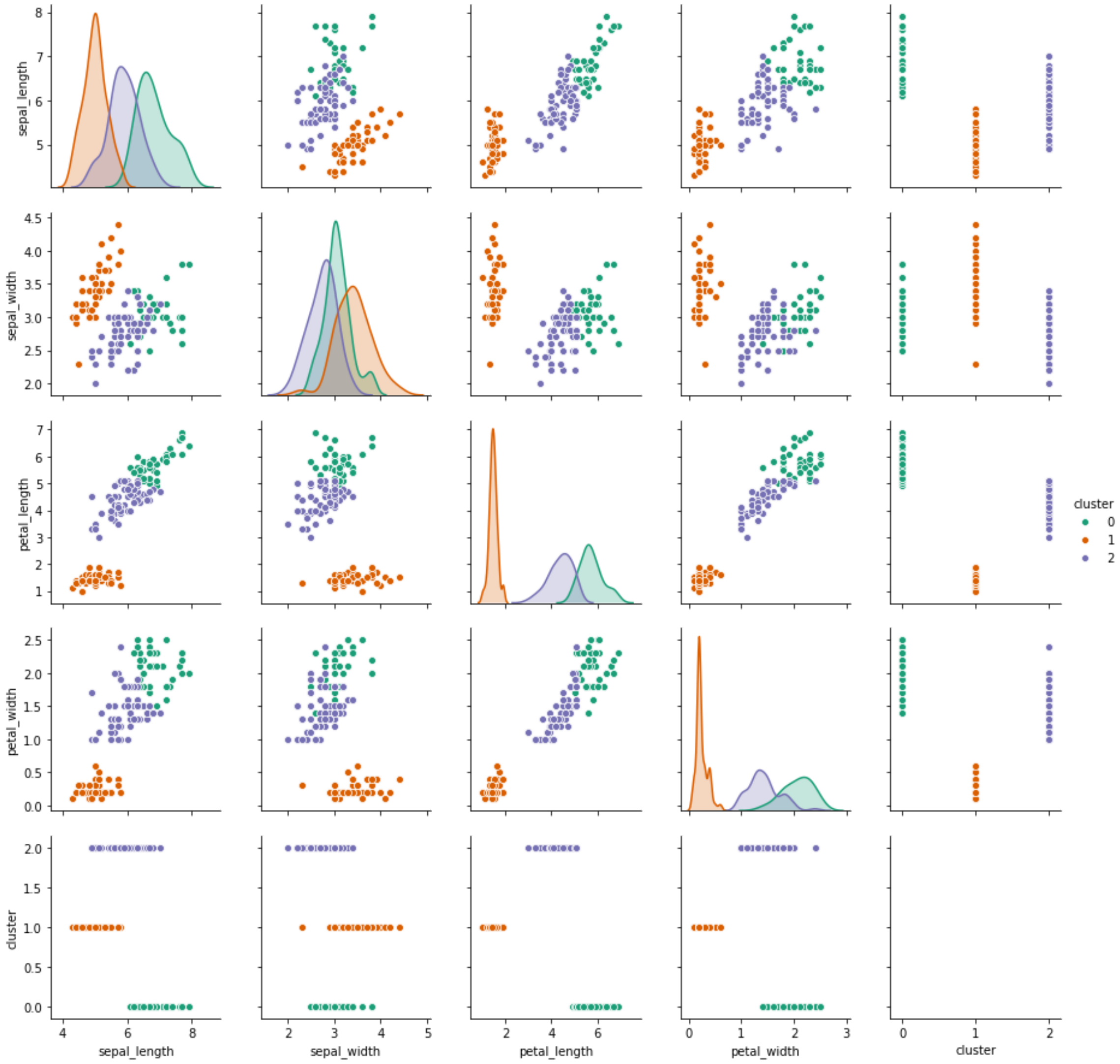
```
df.head(20)
```

Out[81]:

	sepal_length	sepal_width	petal_length	petal_width	species	cluster
0	5.1	3.5	1.4	0.2	setosa	1
1	4.9	3.0	1.4	0.2	setosa	1
2	4.7	3.2	1.3	0.2	setosa	1
3	4.6	3.1	1.5	0.2	setosa	1
4	5.0	3.6	1.4	0.2	setosa	1
5	5.4	3.9	1.7	0.4	setosa	1
6	4.6	3.4	1.4	0.3	setosa	1
7	5.0	3.4	1.5	0.2	setosa	1
8	4.4	2.9	1.4	0.2	setosa	1
9	4.9	3.1	1.5	0.1	setosa	1
10	5.4	3.7	1.5	0.2	setosa	1
11	4.8	3.4	1.6	0.2	setosa	1
12	4.8	3.0	1.4	0.1	setosa	1
13	4.3	3.0	1.1	0.1	setosa	1
14	5.8	4.0	1.2	0.2	setosa	1
15	5.7	4.4	1.5	0.4	setosa	1
16	5.4	3.9	1.3	0.4	setosa	1
17	5.1	3.5	1.4	0.3	setosa	1
18	5.7	3.8	1.7	0.3	setosa	1
19	5.1	3.8	1.5	0.3	setosa	1

```
In [80]: sns.pairplot(df, hue='cluster', palette='Dark2')
         plt.show()
```

```
D:\jupter\lib\site-packages\statsmodels\nonparametric\kde.py:487: RuntimeW
arning: invalid value encountered in true_divide
  binned = fast_linbin(X, a, b, gridsize) / (delta * nobs)
D:\jupter\lib\site-packages\statsmodels\nonparametric\kdetools.py:34: Runt
imeWarning: invalid value encountered in double_scalars
  FAC1 = 2*(np.pi*bw/RANGE)**2
```



```
In [ ]:
```