In [47]:
```
import pandas as pd
import numpy as np
```

In [48]:
```
fd=pd.read_csv('flower_dataset - Copy.csv')
```

In [49]:
```
fd
```

Out[49]:
```
              species    size  fragrance  height_cm  petal_length  petal_width  sepal_length  sepal_width
    0            rose  medium       mild      48.55        2.4275       1.9420        33.985       14.565
    1            rose  medium       mild      48.55        2.4275       1.9420        33.985       14.565
    2 shoeblack plant  medium       mild     147.07        7.3535       5.8828       102.949       44.121
    3 shoeblack plant  medium       none     102.93        5.1465       4.1172        72.051       30.879
    4        hibiscus   large       none     184.00        9.2000       7.3600       128.800       55.200
  ...             ...     ...        ...        ...           ...          ...           ...          ...
 9996        hibiscus   large       none     109.52        5.4760       4.3808        76.664       32.856
 9997 shoeblack plant  medium       mild     145.23        7.2615       5.8092       101.661       43.569
 9998        hibiscus   large       none     126.69        6.3345       5.0676        88.683       38.007
 9999 shoeblack plant   large       none      77.62        3.8810       3.1048        54.334       23.286
10000            rose  medium       mild      88.11        4.4055       3.5244        61.677       26.433

10001 rows × 8 columns
```

In [50]:
```
fd.head()
```

Out[50]:

	species	size	fragrance	height_cm	petal_length	petal_width	sepal_length	sepal_width
0	rose	medium	mild	48.55	2.4275	1.9420	33.985	14.565
1	rose	medium	mild	48.55	2.4275	1.9420	33.985	14.565
2	shoeblack plant	medium	mild	147.07	7.3535	5.8828	102.949	44.121
3	shoeblack plant	medium	none	102.93	5.1465	4.1172	72.051	30.879
4	hibiscus	large	none	184.00	9.2000	7.3600	128.800	55.200

In [51]:

```
fd.tail()
```

Out[51]:

	species	size	fragrance	height_cm	petal_length	petal_width	sepal_length	sepal_width
9996	hibiscus	large	none	109.52	5.4760	4.3808	76.664	32.856
9997	shoeblack plant	medium	mild	145.23	7.2615	5.8092	101.661	43.569
9998	hibiscus	large	none	126.69	6.3345	5.0676	88.683	38.007
9999	shoeblack plant	large	none	77.62	3.8810	3.1048	54.334	23.286
10000	rose	medium	mild	88.11	4.4055	3.5244	61.677	26.433

In [52]:

```
fd.describe()
```

Out[52]:

	height_cm	petal_length	petal_width	sepal_length	sepal_width
count	10001.000000	10001.000000	10001.000000	10001.000000	10001.000000
mean	104.868306	5.243415	4.194732	73.407814	31.460492
std	43.636644	2.181832	1.745466	30.545651	13.090993
min	30.090000	1.504500	1.203600	21.063000	9.027000
25%	69.300000	3.465000	2.772000	48.510000	20.790000
50%	100.500000	5.025000	4.020000	70.350000	30.150000
75%	137.020000	6.851000	5.480800	95.914000	41.106000
max	199.970000	9.998500	7.998800	139.979000	59.991000

In [53]:

```
fd.isna().sum()
```

Out[53]:

```
species         0
size            0
fragrance       0
height_cm       0
petal_length    0
petal_width     0
sepal_length    0
sepal_width     0
dtype: int64
```

In [54]:

```
from sklearn.preprocessing import LabelEncoder

# Initialize the label encoder
label_encoder = LabelEncoder()
fd['species']=label_encoder.fit_transform(fd['species'])
fd['size']=label_encoder.fit_transform(fd['size'])
fd['fragrance']=label_encoder.fit_transform(fd['fragrance'])
```

In [55]:

```
size_vale=fd['size']
size_vale
```

Out[55]:

```
0        1
1        1
2        1
3        1
4        0
        ..
9996     0
9997     1
9998     0
9999     0
10000    1
Name: size, Length: 10001, dtype: int32
```

In [56]:

```
fd.corr()
```

Out[56]:

	species	size	fragrance	height_cm	petal_length	petal_width	sepal_length	sepal_width
species	1.000000	0.273530	-0.294099	-0.453949	-0.453949	-0.453949	-0.453949	-0.453949
size	0.273530	1.000000	0.046752	-0.638693	-0.638693	-0.638693	-0.638693	-0.638693
fragrance	-0.294099	0.046752	1.000000	0.025577	0.025577	0.025577	0.025577	0.025577
height_cm	-0.453949	-0.638693	0.025577	1.000000	1.000000	1.000000	1.000000	1.000000
petal_length	-0.453949	-0.638693	0.025577	1.000000	1.000000	1.000000	1.000000	1.000000
petal_width	-0.453949	-0.638693	0.025577	1.000000	1.000000	1.000000	1.000000	1.000000
sepal_length	-0.453949	-0.638693	0.025577	1.000000	1.000000	1.000000	1.000000	1.000000
sepal_width	-0.453949	-0.638693	0.025577	1.000000	1.000000	1.000000	1.000000	1.000000

In [57]:

```
correlation=fd.corr()['species']
correlation
```

Out[57]:

```
species         1.000000
size            0.273530
fragrance      -0.294099
height_cm      -0.453949
petal_length   -0.453949
petal_width    -0.453949
sepal_length   -0.453949
sepal_width    -0.453949
Name: species, dtype: float64
```

In [58]:

```
x=fd.drop(['size','fragrance','species'],axis=1)
x
```

Out[58]:

	height_cm	petal_length	petal_width	sepal_length	sepal_width
0	48.55	2.4275	1.9420	33.985	14.565
1	48.55	2.4275	1.9420	33.985	14.565
2	147.07	7.3535	5.8828	102.949	44.121
3	102.93	5.1465	4.1172	72.051	30.879
4	184.00	9.2000	7.3600	128.800	55.200
...	...	...	...	...	...
9996	109.52	5.4760	4.3808	76.664	32.856
9997	145.23	7.2615	5.8092	101.661	43.569
9998	126.69	6.3345	5.0676	88.683	38.007
9999	77.62	3.8810	3.1048	54.334	23.286
10000	88.11	4.4055	3.5244	61.677	26.433

10001 rows × 5 columns

In [59]:

```
y=fd['species']
y
```

Out[59]:
```
0        1
1        1
2        2
3        2
4        0
        ..
9996     0
9997     2
9998     0
9999     2
10000    1
Name: species, Length: 10001, dtype: int32
```

In [60]:
```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 42)
```

In [61]:
```
X_train.shape
```

Out[61]:
```
(8000, 5)
```

In [62]:
```
y_train.shape
```

Out[62]:
```
(8000,)
```

In [63]:
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score
RFC = RandomForestClassifier()
```

In [64]:
```
RFC.fit(X_train,y_train)
```

Out[64]:
```
▾ RandomForestClassifier
RandomForestClassifier()
```

In [65]:
```
test_predict=RFC.predict(X_test)
RAccuracy=Accuracy=accuracy_score(y_test,test_predict)
```

```
In [66]: RAccuracy*100
```

```
Out[66]: 64.01799100449776
```

```
In [67]: import pickle
```

```
In [68]: with open('rose_1.pkl', 'wb') as file:
             pickle.dump(RFC, file)
```

```
In [ ]:
```