In [1]:

```python
import pandas as pd
import seaborn as sns
penguins = sns.load_dataset("penguins")
```

In [2]:

```python
penguins.head()
```

Out[2]:

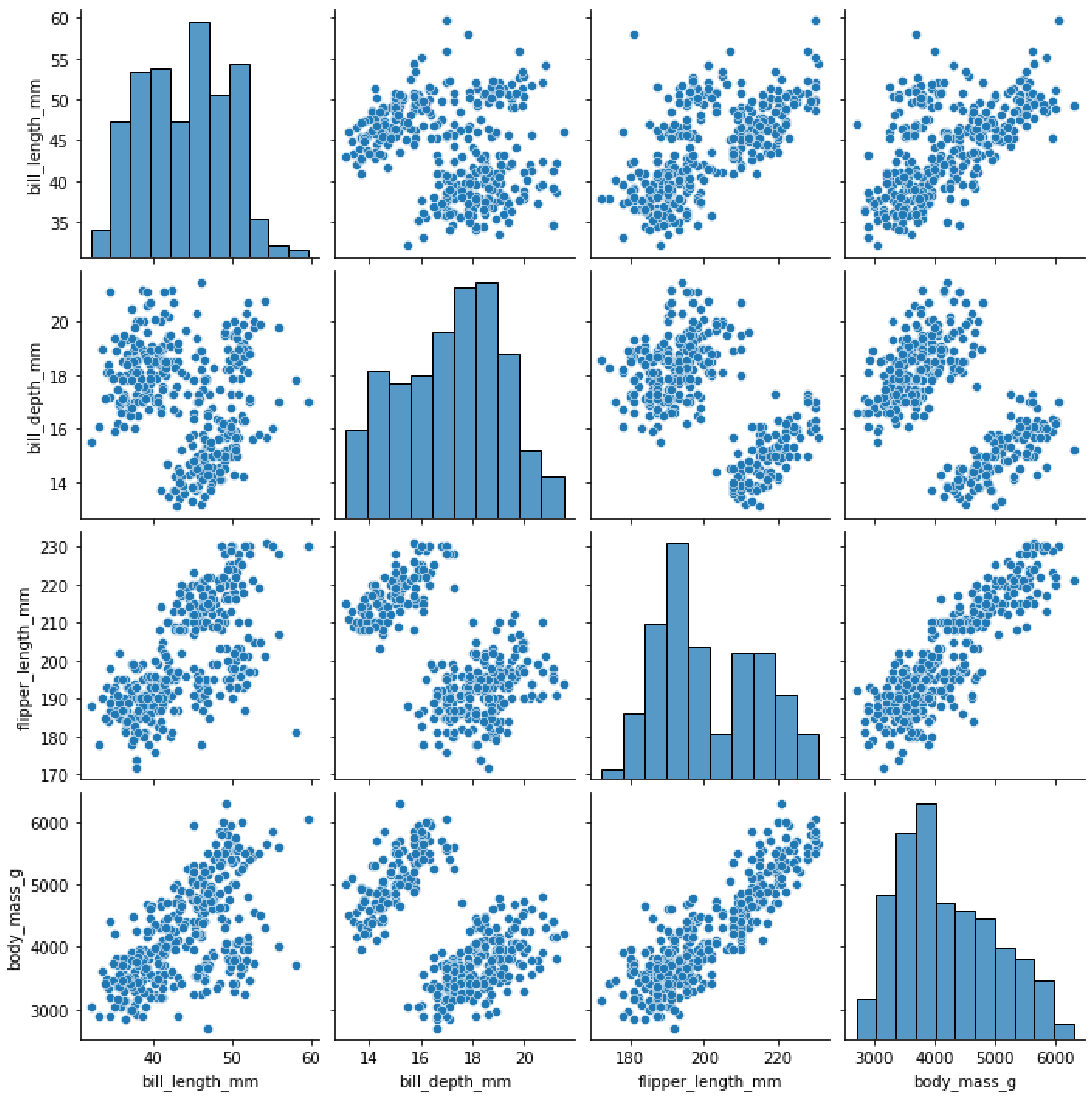
	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female

In [4]:

```python
sns.pairplot(penguins)
```

Out[4]:

```
<seaborn.axisgrid.PairGrid at 0x1e2ee31dd00>
```



In each graph clusters can be found. the biggest cluster are the top right and bottom left.

In [7]:

```python
from sklearn.cluster import KMeans
```

In [8]:

```python
penguinsNan = penguins.dropna()
```

In [9]:

```python
features = ['bill_length_mm','bill_depth_mm','flipper_length_mm', 'body_mass_g']
km = KMeans(n_clusters=2, random_state=42).fit(penguinsNan[features])
```

In [13]:

```python
penguinsNan.head()
```

Out[13]:

	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	sex	cluster
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	Male	0
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	Female	0
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	Female	0
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	Female	0
5	Adelie	Torgersen	39.3	20.6	190.0	3650.0	Male	0

In [14]:

```python
from sklearn import metrics
from sklearn.metrics import pairwise_distances
```

In [16]:

```python
metrics.silhouette_score(penguinsNan[features], km.labels_, metric='euclidean')
```

Out[16]:

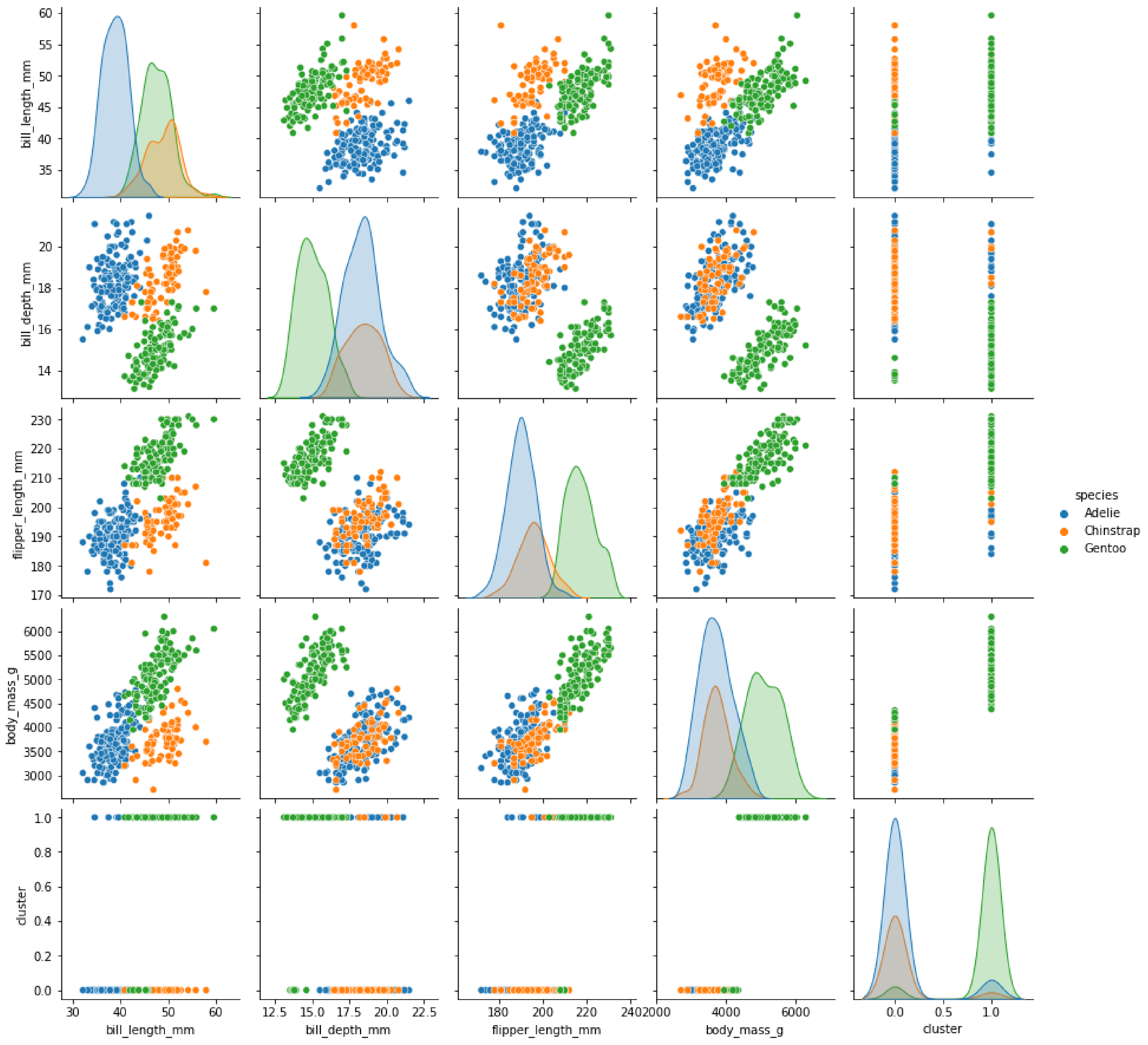
```
0.6307117469850305
```

In [17]:

```python
sns.pairplot(penguinsNan, hue="species")
```

Out[17]:

```
<seaborn.axisgrid.PairGrid at 0x1e2ef4e9d90>
```



In [18]:

```python
contingency_table = penguinsNan.groupby(['species','cluster']).size().unstack('cluster', fill_value=0)
contingency_table
```

Out[18]:

cluster	0	1
species		
Adelie	132	14
Chinstrap	63	5
Gentoo	8	111

There were no exact matches