

```
In [1]: import pandas as pd
import seaborn as sns
penguins = sns.load_dataset("penguins")
```

```
In [2]: penguins.head()
```

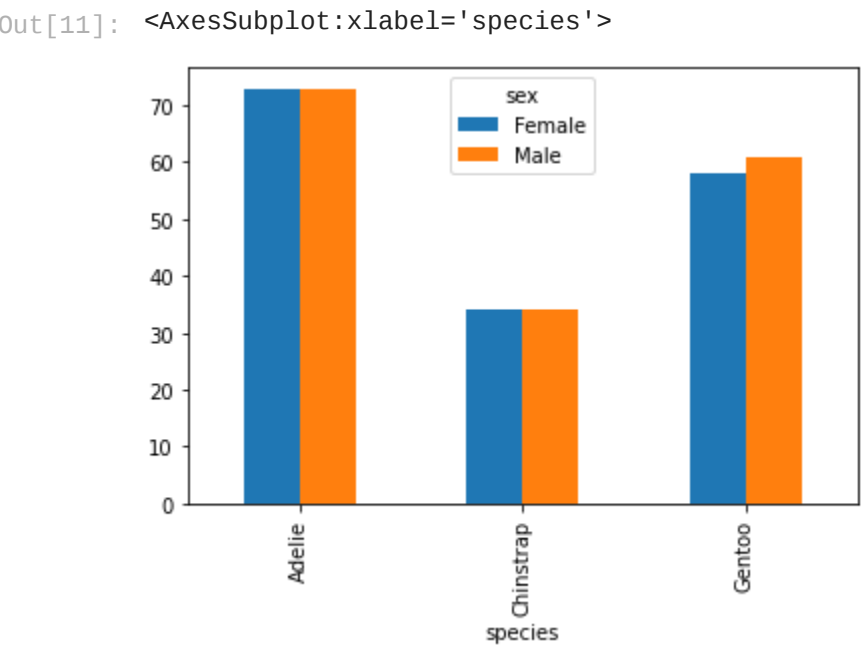
Out[2]:

	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

Expectations

I believe the columns will correlate.

```
In [11]: penguins.groupby(['species', 'sex']).size().unstack('sex', fill_value=0).plot(kind='bar')
```



The amount of penguins of each sex seems to be around the same with each of the 3 species.

```
In [12]: penguins.groupby(['species', 'sex']).size().unstack('sex', fill_value=0)
```

Out[12]:

species	sex	Female	Male
Adelie		73	73
Chinstrap		34	34
Gentoo		58	61

```
In [17]: from scipy.stats import chi2_contingency
```

```
In [18]: chi2_contingency(penguins.groupby(['species', 'sex']).size().unstack('sex', fill_value=0))
```

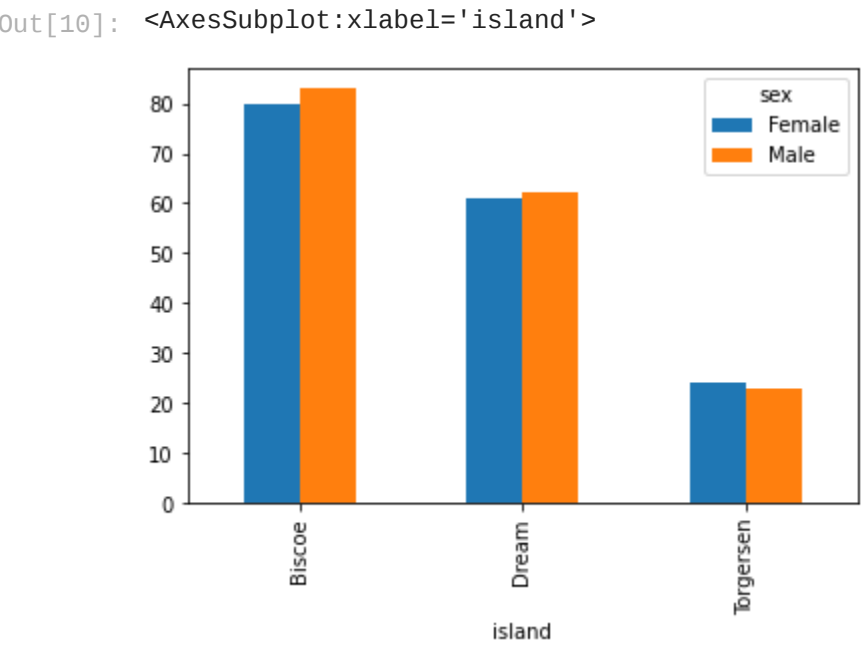
Out[18]: (0.04860717014078318,  
0.9759893689765846,  
2,  
array([[72.34234234, 73.65765766],  
[33.69369369, 34.30630631],  
[58.96396396, 60.03603604]]))

There seems to be a 98% chance the columns aren't the same.

Expectations

I believe the columns will correlate.

```
In [10]: penguins.groupby(['island', 'sex']).size().unstack('sex', fill_value=0).plot(kind='bar')
```



The amount of penguins of each sex seems to be around the same on each of the 3 islands.

```
In [19]: penguins.groupby(['island', 'sex']).size().unstack('sex', fill_value=0)
```

Out[19]:

island	sex	Female	Male
Biscoe		80	83
Dream		61	62
Torgersen		24	23

```
In [20]: chi2_contingency(penguins.groupby(['island', 'sex']).size().unstack('sex', fill_value=0))
```

Out[20]: (0.05759904881286207,  
0.971611229281065,  
2,  
array([[80.76576577, 82.23423423],  
[60.94594595, 62.05405405],  
[23.28828829, 23.71171171]]))

There seems to be a 97% chance the columns aren't the same.