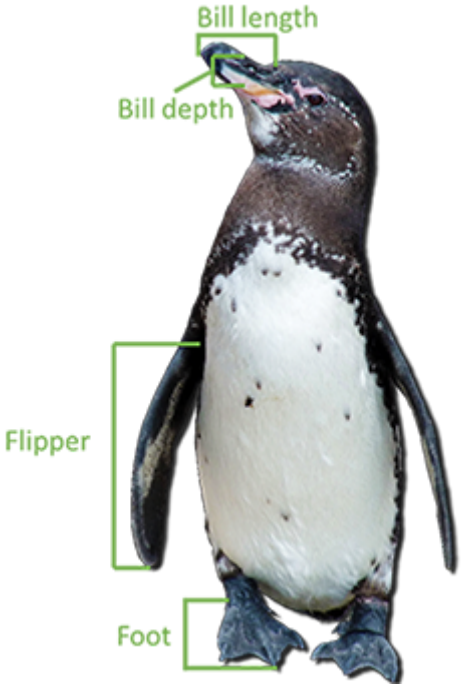
## Portfolio assignment 11

20 min: Do a Numerical VS Categorical bivariate analysis on the penguins dataset.

- Choose one of the categorical columns: species, island or sex
- use .groupby('').mean() too look at the means of the numerical columns. Does it look like there is a difference between categories?
- Use the seaborn barplot to plot the mean and confidence. Create this plot for each of the numerical columns (bill_length_mm bill_depth_mm, flipper_length_mm, body_mass_g)
- For each of the plots, write a conclusion: Is there a statistically significant difference for this numerical column for each category?
- Optional: Repeat this proces for the other two categorical columns




In [1]:

```
import seaborn as sns
```

In [2]:

```
penguins = sns.load_dataset("penguins")
penguins.head()
```

Out[2]:

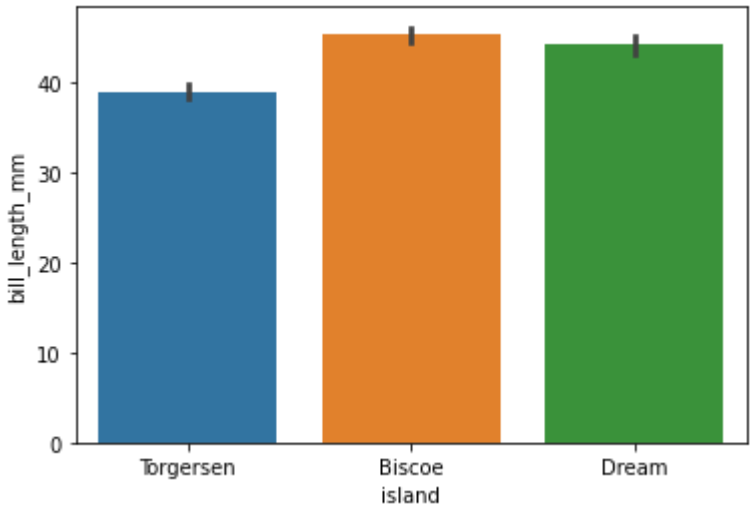
| | species | island | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g | sex |
|---|---|---|---|---|---|---|---|
| **0** | Adelie | Torgersen | 39.1 | 18.7 | 181.0 | 3750.0 | Male |
| **1** | Adelie | Torgersen | 39.5 | 17.4 | 186.0 | 3800.0 | Female |
| **2** | Adelie | Torgersen | 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 [3]:

```
sns.barplot(y="bill_length_mm", x="island", data=penguins)
penguins.groupby(by="island").mean()
```

Out[3]:

| island | bill_length_mm | bill_depth_mm | flipper_length_mm | body_mass_g |
|---|---|---|---|---|
| **Biscoe** | 45.257485 | 15.874850 | 209.706587 | 4716.017964 |
| **Dream** | 44.167742 | 18.344355 | 193.072581 | 3712.903226 |
| **Torgersen** | 38.950980 | 18.429412 | 191.196078 | 3706.372549 |



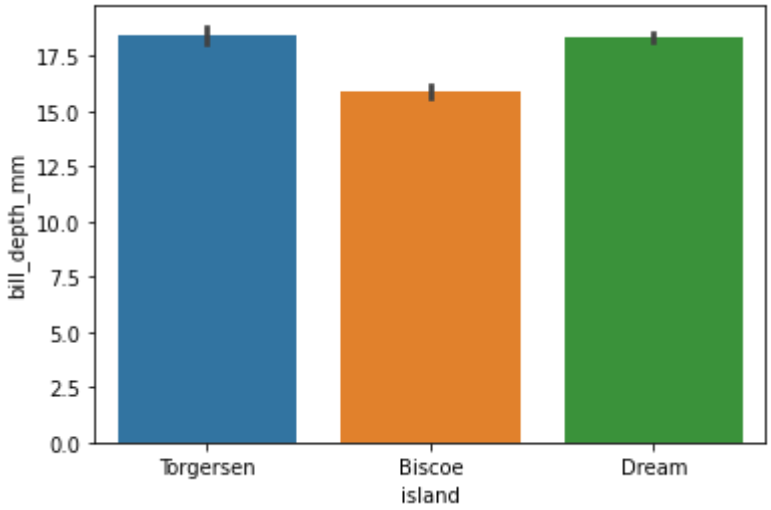

There is a significant difference between the Torgersen island and the other two, you can make the affirmation that Pinguins who have a Bil length of less than 4 lives on the Torgersen island. It might have something to do with the species that live there and their caracteristics.

In [4]:

```
sns.barplot(y="bill_depth_mm", x="island", data=penguins)
```

Out[4]:

```
<AxesSubplot:xlabel='island', ylabel='bill_depth_mm'>
```



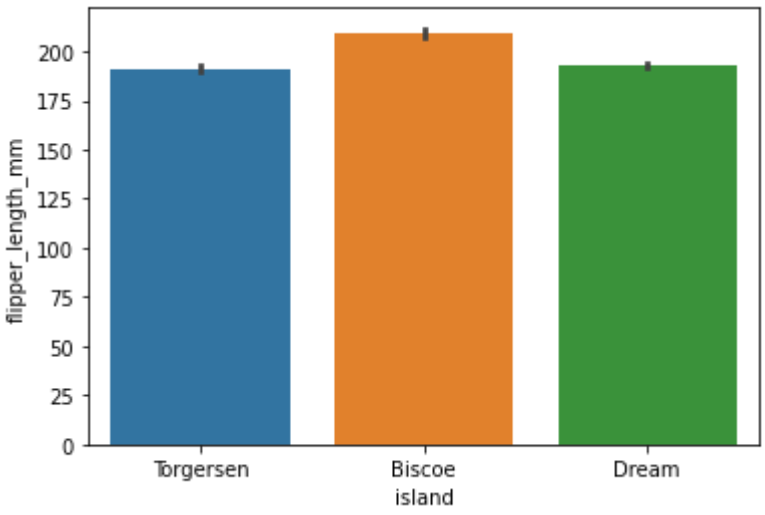

There can be noticed that pinguins that have a Bil depth less than 16 live in majority on the Biscoe Island, it can have something to do with the specific species that live on the island. The rest of the pinguins are divided on the other two islands.

In [5]:

```
sns.barplot(y="flipper_length_mm", x="island", data=penguins)
```

Out[5]:

```
<AxesSubplot:xlabel='island', ylabel='flipper_length_mm'>
```



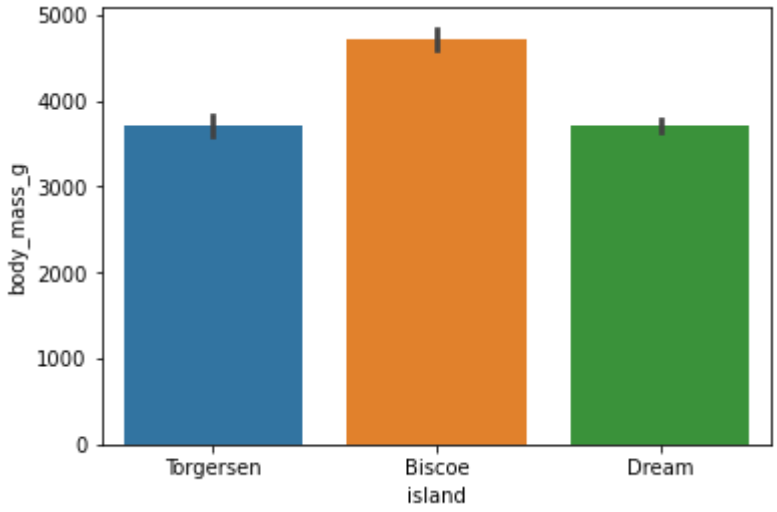

The Biscoe Island seems to have more pinguins with a biggfer Flipper length, again it might have something to do with the species.

In [6]:

```
sns.barplot(y="body_mass_g", x="island", data=penguins)
```

Out[6]:

```
<AxesSubplot:xlabel='island', ylabel='body_mass_g'>
```




The body mass of the pinguins are higher at the Biscoe Islad, and the other two seem to have a similar distribuition.