

**TASK**

**Exploratory Data Analysis on the penguins\_size Dataset**

[](http://www.hyperiondev.com/portal/)

**Introduction**

This dataset contains measurements for penguins, different species, the island they inhabit, culmen length and depth in, flipper length, body mass and sex. It allows us to study the physical attributes of penguins with their unique species and islands they inhabit. We can explore the relationships of their physical features between males and females.

**DATA CLEANING**

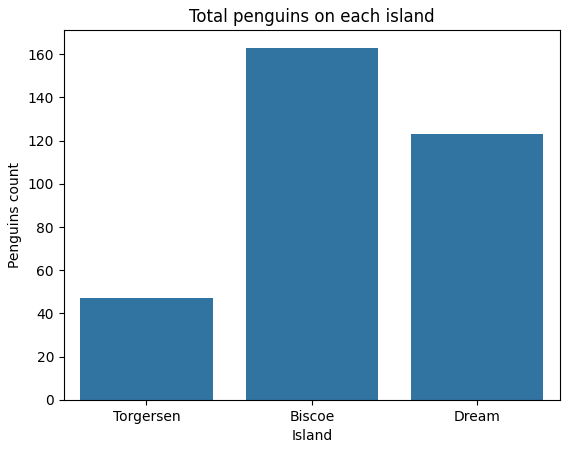
Checked each column’s data type then change flipper length and body mass data types to NumPy’s int64 data type.

Removed any duplicate rows in the dataset.

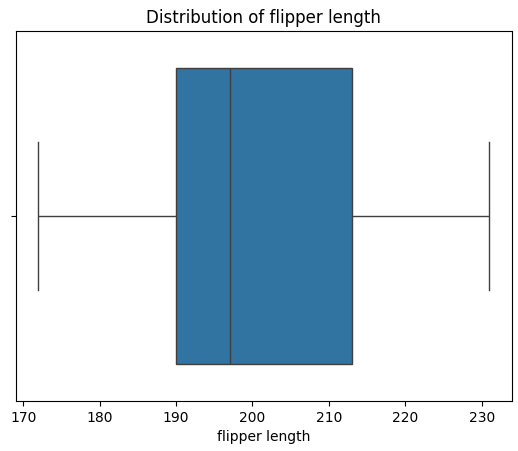
**MISSING DATA**

Removed any rows with Na values and the row with a "." value for sex.

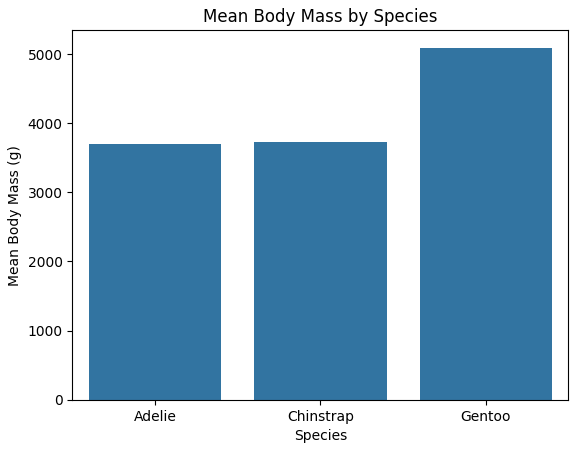
**DATA STORIES AND VISUALISATIONS**



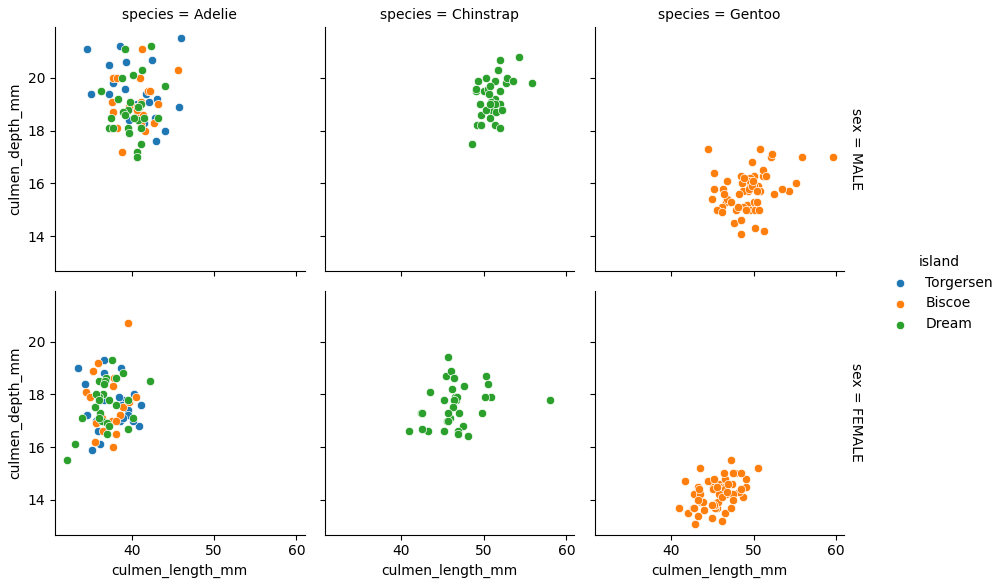
From this dataset we can see that Biscoe is the most populated Island for penguins at around 160 penguins followed by Dream and then Torgersen with roughly 120 and 50 penguins.



This box plot tells us that most of penguin's flipper length are between the lengths of roughly 190 and 215 mm.



We can see that the species with the meanest body mass is by far Gentoo penguins with around 5000 grams average while Adelie and Chinstrap penguins have roughly the same average body mass with just under 4000 grams.



This multigrid tells us that for most of the penguins, males have bigger culmen lengths and depths than females do. We can see that Adelie penguins inhabit all three islands including Torgersen-, Biscoe- and Dream Island while Chinstrap penguins only inhabit Dream Island and Gentoo penguins only inhabit Biscoe island. For all three penguin species, there is a positive relationship between culmen length and culmen depth.

**THIS REPORT WAS WRITTEN BY: Christiaan Cronje**

