



TASK

Exploratory Data Analysis on the Penguins_Size Dataset

[Visit our website](#)

Introduction

This is a Data Visualisation Capstone Project which will be dealing with Penguins_Size Dataset, it specifically looks at the three species of penguins (Adelie, Gentoo and Chinstrap) collected from various islands (Biscoe, Dream and Torgersen). The dataset includes features such as species, island, culmen length (mm), culmen depth (mm), flipper length (mm), body mass (g) and sex.

In this document I will be sharing how I have analysed this data set using the knowledge I have gained in data cleaning, handling missing data and data visualisations.

DATA CLEANING AND MISSING DATA

I decided to combine data cleaning and missing data in this Capstone Project because in my solution I found a way to do them simultaneously.

I first checked missing values in the dataset by using `isnull().sum()`, which helped in checking from each column how many missing values there are in that specific column. Almost all the columns consists of missing values except for the columns 'species' and 'island'.

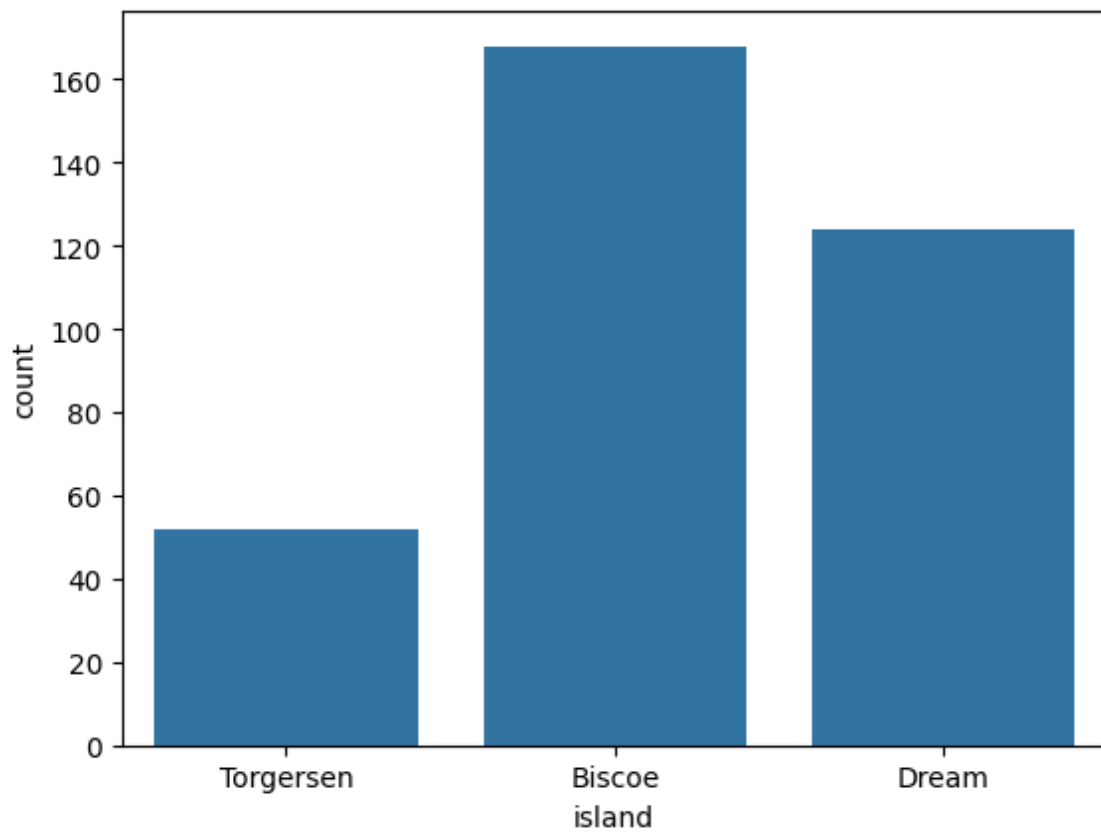
After finding missing values, I then imported SimpleImputer library to fill in the missing values in the datasets. I utilised the 'strategy' to replace the missing data with the 'most frequently' seen value in that specific column.

DATA STORIES AND VISUALISATIONS

Below are interesting visualisations that I have generated.

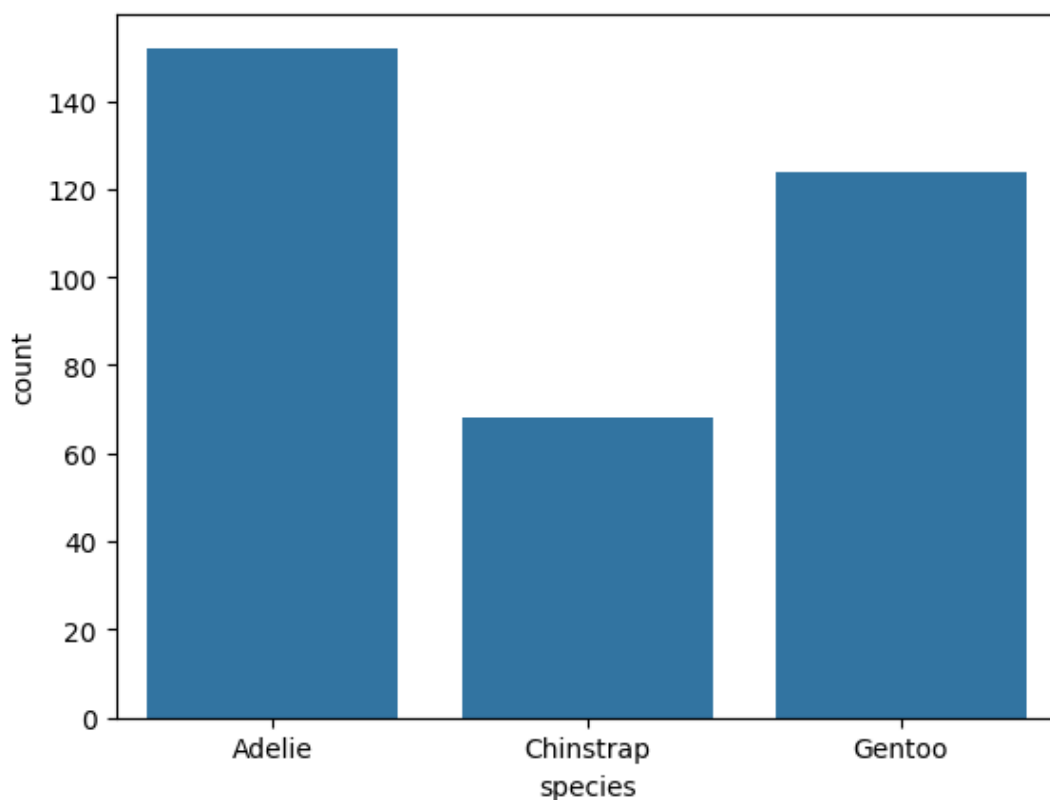
Penguins Per Species

In this plot I counted the number of penguins per species, I found that there are more Adelies, followed by the Gentoos and the least in number are Chinstraps.



Penguins per island

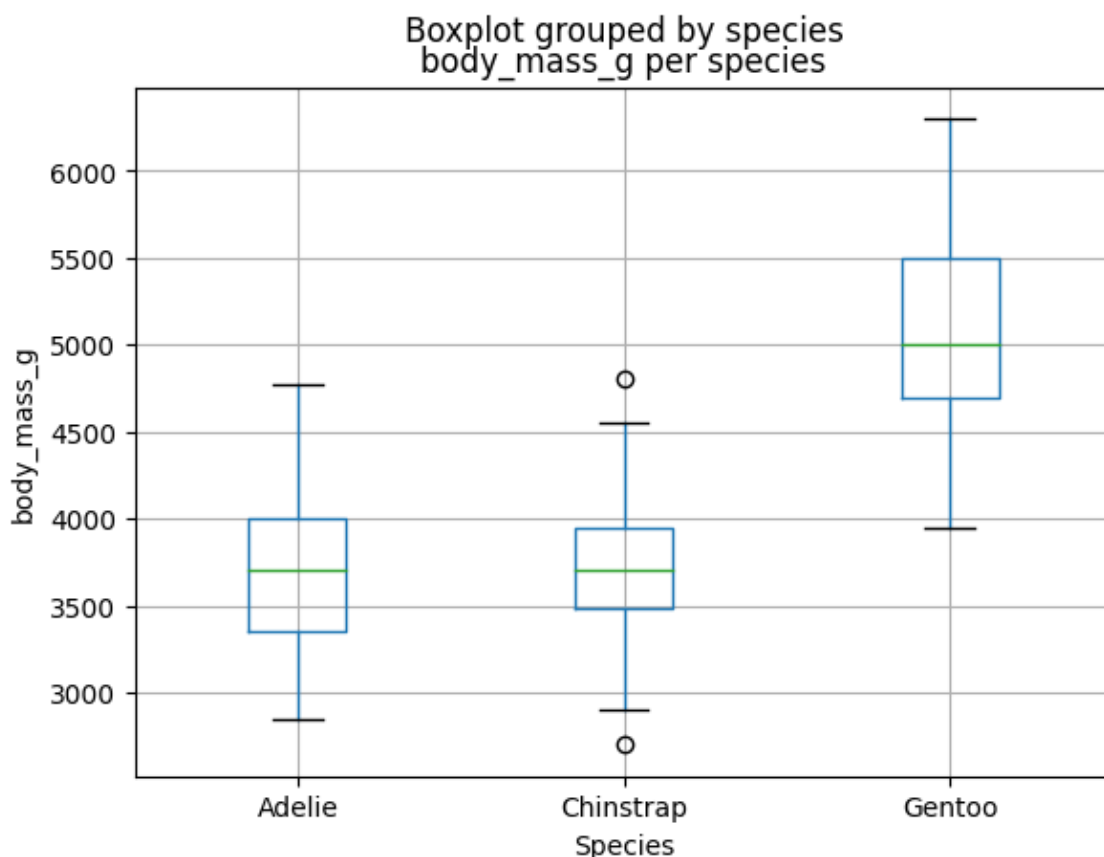
After getting how many penguins there are in each species, I was interested to know the number of penguins there are per island. I followed exactly the same way I did above and the box plot came out as shown below.



It is found that there are more penguins in Biscoe Island, followed by Dream Island and the island with the least penguins is Torgersen.

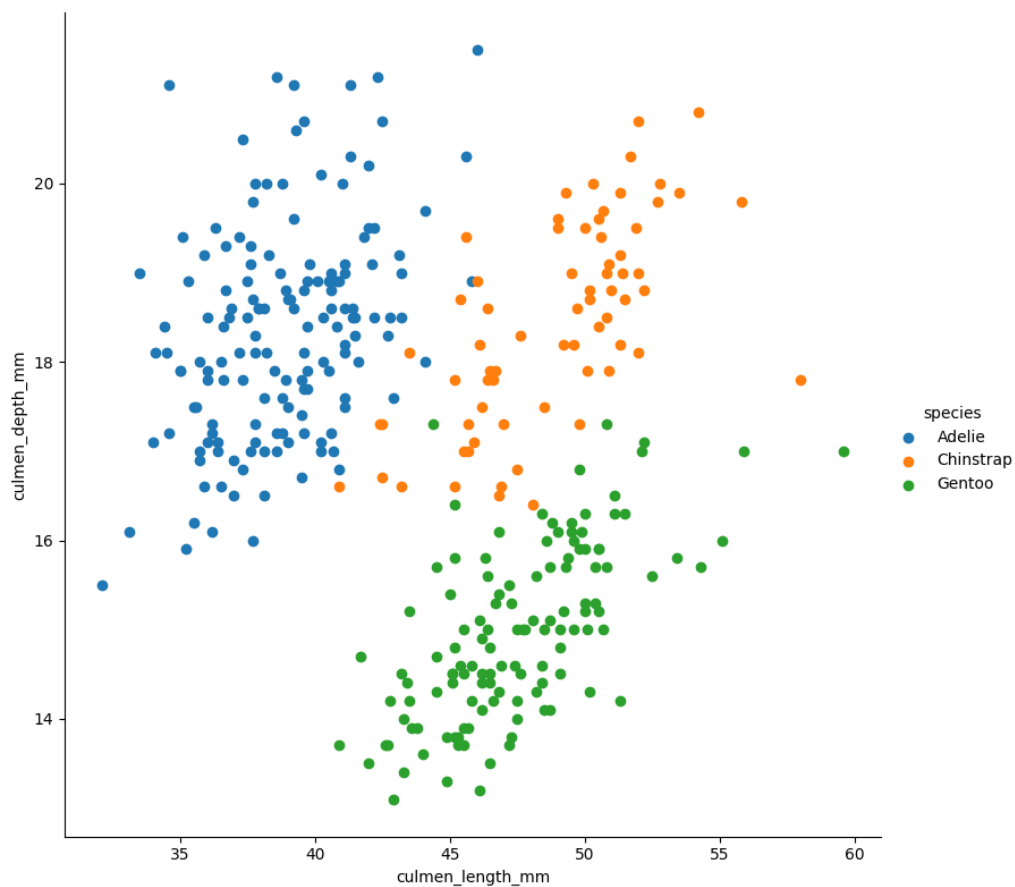
Body mass per species

I used box plot to generate this graph. I was interested to see which of the three species have heavier penguins, I found that the Gentoo are the heaviest in mass, followed by the Adelie which weigh slightly more than the Chinstrap.



Culmen depth vs culmen length

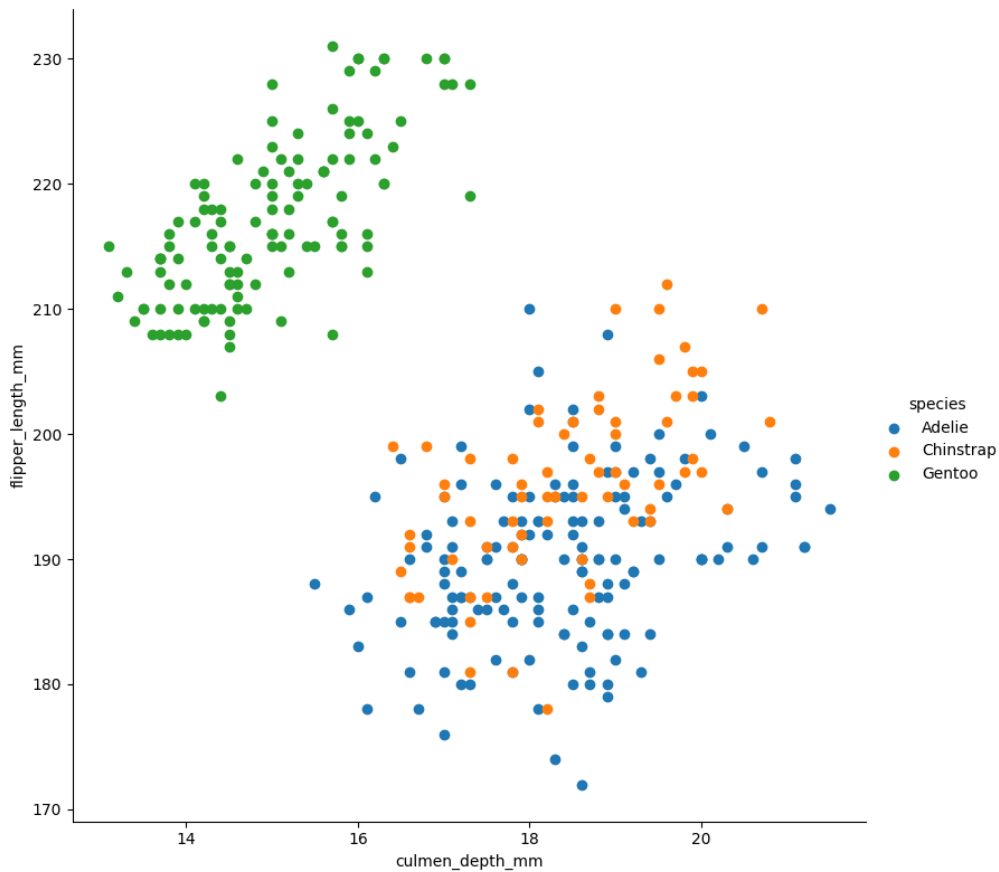
I used a scatter plot to compare the culmen depth and length of the three species. The Adelie species was found to have the lowest culmen length, while the Chinstrap and Gentoo have about the same culmen length. Coming to culmen depth, Adelie and Chinstrap have about the same depth while Gentoo have shorter length.



Culmen depth vs flipper length

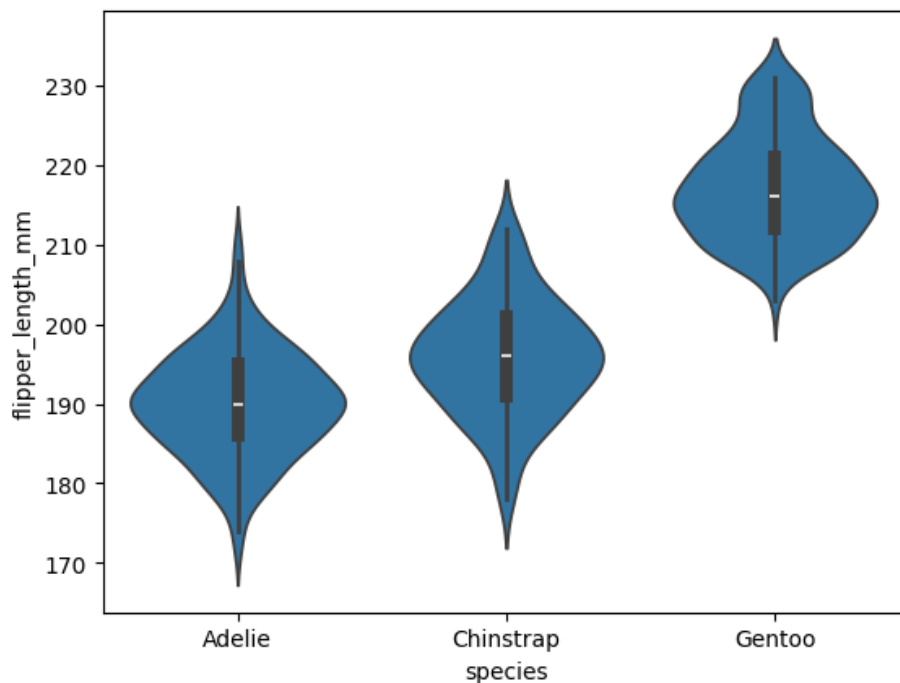
In this visualisation I am comparing culmen depth and flipper length of the three penguin species. I found that Chinstrap and Adelie culmen depth is almost similar, but the flipper length differs with that of Chinstrap slightly bigger than that of Adelie.

On the other hand Gentoo have smaller culmen depth and bigger flipper length than the other two species.



Flipper length distribution

I used a violin plot to clearly show how the flipper lengths of the three species differ. Gentoo flipper length is the biggest ranging from approximately 200mm to 235mm, it is seen that the most Gentoos flipper length is around 215mm. Adelie flipper length ranges from approximately 165mm to 205mm, it is seen that there are more Adelie penguins with flipper length of approximately 190mm and seems to be the lowest one out of the three species.



THIS REPORT WAS WRITTEN BY : NONOPHA GEGE
