Applied Data Science Assignment 2

General analysis of penguin characteristics

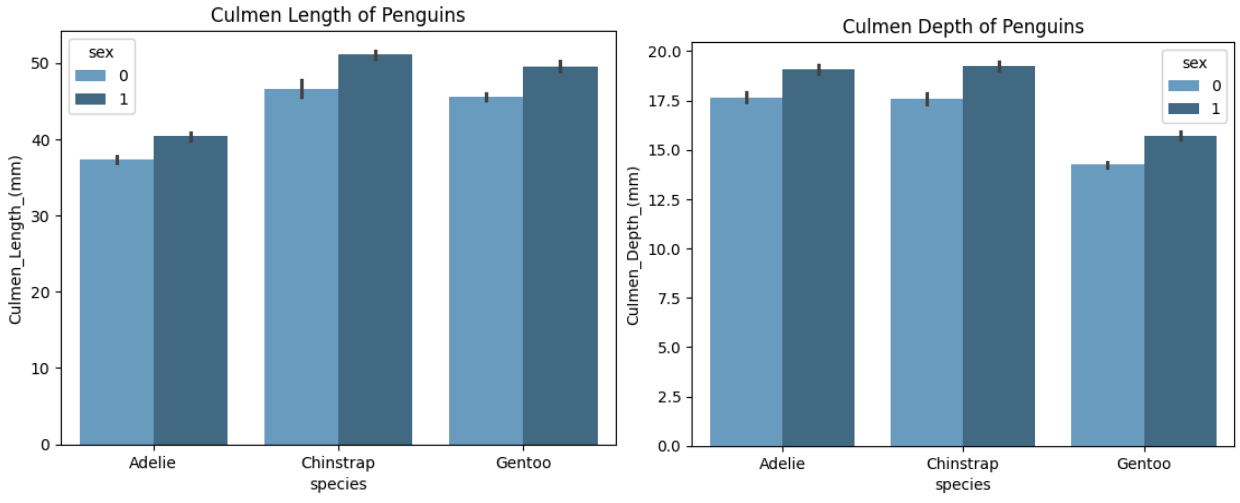
Src: https://www.kaggle.com/datasets/amulyas/penguin-size-dataset?resource=download
https://github.com/leesiro12/Applied-Data-Science-Assignment-2

This is an analysis report of general characteristics of 3 species of penguins with more emphasis on their culmen length and depth. The table below has been formatted to allow easier viewing and processing.

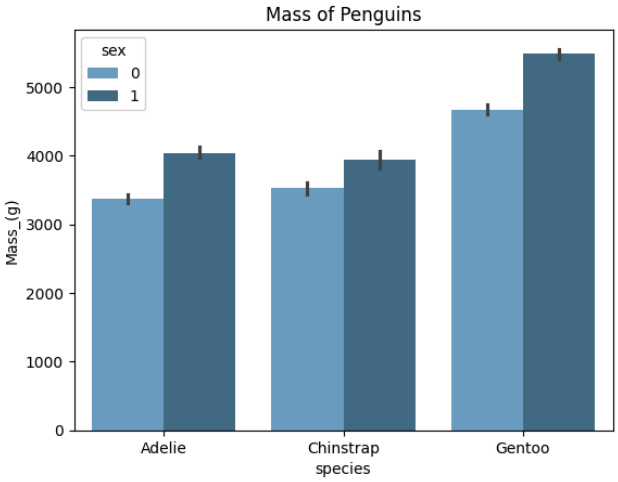
(*0=female; 1=male)

	species	Culmen_Length_(mm)	Culmen_Depth_(mm)	Flipper_Length_(mm)	Mass_(g)	sex
0	Adelie	39.1	18.7	181.0	3750.0	1
1	Adelie	39.5	17.4	186.0	3800.0	0
2	Adelie	40.3	18.0	195.0	3250.0	0
4	Adelie	36.7	19.3	193.0	3450.0	0
5	Adelie	39.3	20.6	190.0	3650.0	1
...	...	...	...	...	...	...
338	Gentoo	47.2	13.7	214.0	4925.0	0
340	Gentoo	46.8	14.3	215.0	4850.0	0
341	Gentoo	50.4	15.7	222.0	5750.0	1
342	Gentoo	45.2	14.8	212.0	5200.0	0
343	Gentoo	49.9	16.1	213.0	5400.0	1

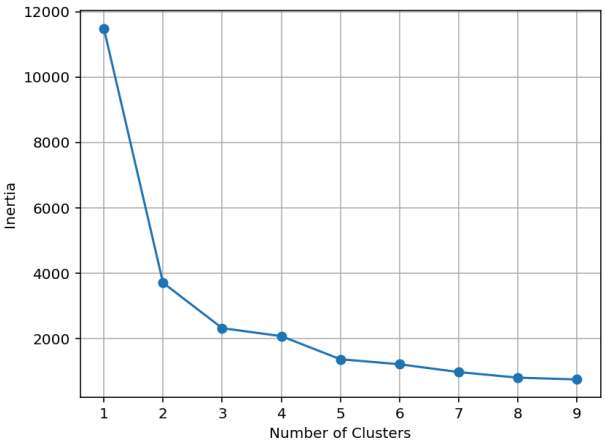
The bar charts below compare and show differences between these penguins. Data shows that males have longer and higher culmen with the Chinstraps at the top.

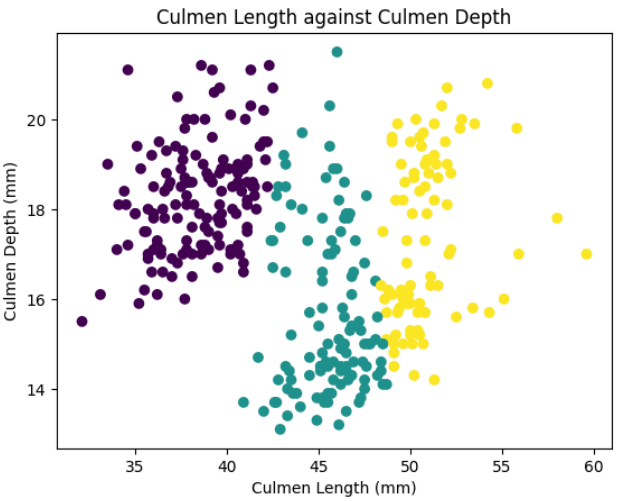


The Chinstraps are not the biggest species however with the Gentoos outweighing them by roughly 1200g.



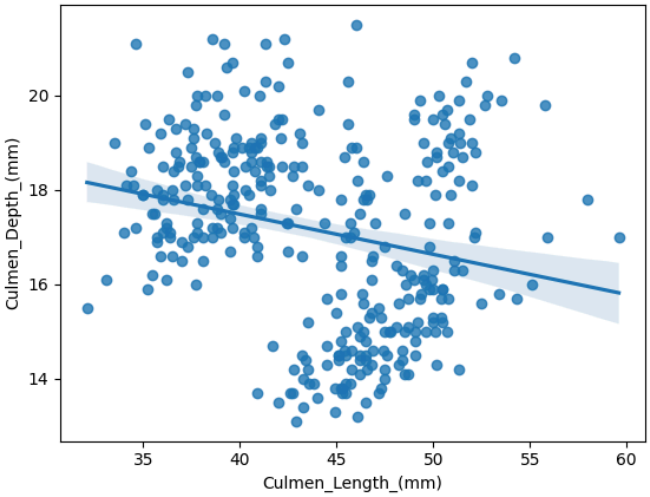
A scatter plot of 3 clusters is generated with depth against length. The elbow method is performed to find the optimal cluster which is 3-4 as seen below. However, using silhouette score gave me a different answer of 2 clusters.



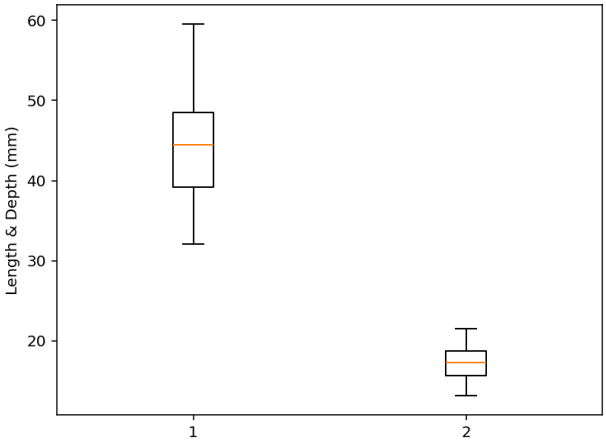


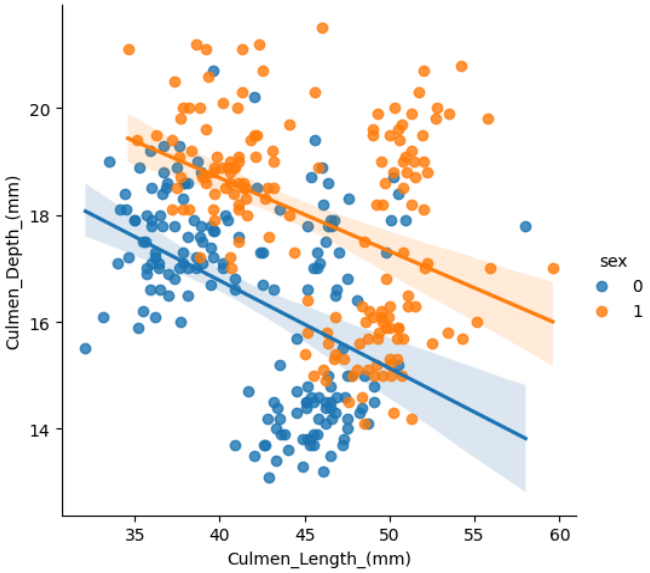

```
 2 clusters silhoutte score = 0.57
 3 clusters silhoutte score = 0.49
 4 clusters silhoutte score = 0.44
 5 clusters silhoutte score = 0.41
 6 clusters silhoutte score = 0.43
 7 clusters silhoutte score = 0.45
 8 clusters silhoutte score = 0.42
 9 clusters silhoutte score = 0.42
10 clusters silhoutte score = 0.41
Best number of clusters =  2
```

Using the silhouette score as the new reference, regression lines can be drawn to fit the data. From the graphs, culmen depth tends to be higher when culmen length is shorter. The sex of the penguins does not seem to influence this trend.



When graphed into a boxplot, the whiskers of both boxes appear to be quite long. Cross referencing between graphs indicates that though there exists a mean size, some penguins are just larger and have bigger bills potentially making them outliers as the average measurements for their species.





Summary:

- Male penguins are larger than female penguins (at least for these species)
- Chinstrap penguins have overall bigger bills
- Gentoo penguins are heavier and/or larger
- The average culmen length is 46mm and the average culmen depth is 17mm for all penguins