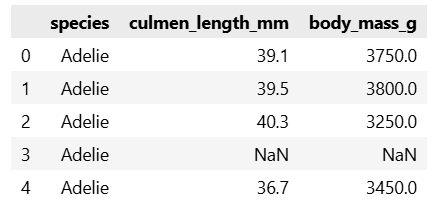
**1.3.2. Bài làm mẫu**

**Bài toán 1:** Thực hiện các nhiệm vụ trong bài toán 2 để làm quen với việc phân tích hai biến với các hàm trong thư viện scikit-learn

**Nhiệm vụ 1:** Phân tích dữ liệu đơn biến trên dữ liệu về chim cánh cụt

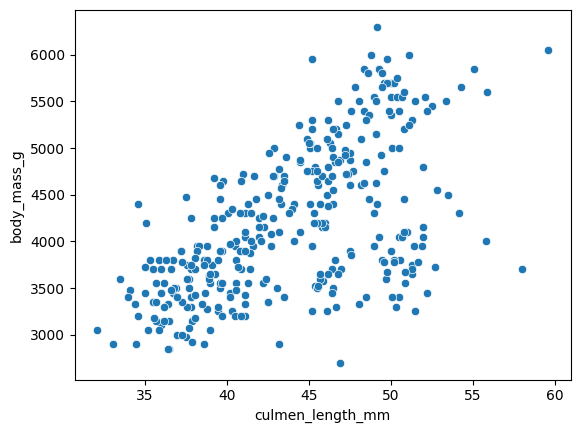
1. Import thư viện và chuẩn bị dữ liệu phân tích

Bảng 1: Dữ liệu Độ dài mỏ và khối lượng cơ thể



Nhận xét: Khi chiều dài mỏ tăng, khối lượng cơ thể chim cánh cụt thường cũng có xu hướng tăng.

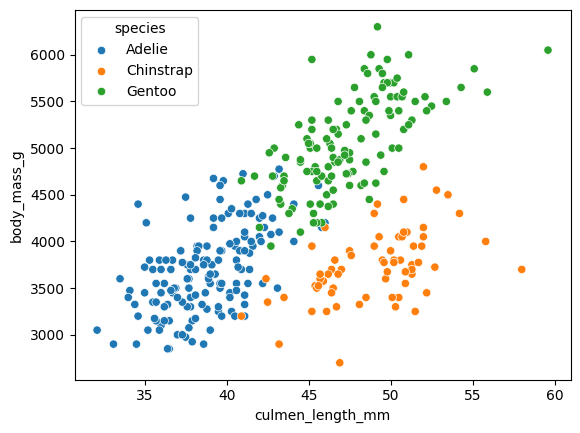
2. Phân tích dữ liệu 2 biến dựa vào phương pháp scatterplot



Hình 1: Biểu đồ phân tán giữa độ dài mỏ và khối lượng cơ thể

Dựa vào hình ta thấy:

Cho thấy mối tương quan dương giữa độ dài mỏ và khối lượng cơ thể.  
Chim cánh cụt có mỏ dài thường nặng hơn.



Hình 2: Biểu đồ phân tán giữa độ dài mỏ và khối lượng cơ thể theo loài

Dựa vào hình ta thấy:

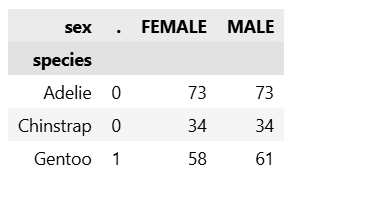
Loài Gentoo có xu hướng mỏ dài hơn và khối lượng cơ thể lớn hơn, cho thấy mối quan hệ tuyến tính khá mạnh.

Loài Adelie có mỏ ngắn và khối lượng nhẹ hơn, nhưng vẫn có sự phân bố tương đối đều theo xu hướng tăng.

Loài Chinstrap nằm giữa hai loài còn lại, với sự phân bố khá chặt chẽ, cho thấy độ tương quan cao giữa hai đặc tính.

3. Phân tích 2 biến dựa vào bảng crosstab/two-way

Bảng 2: Phân bố giới tính theo loài



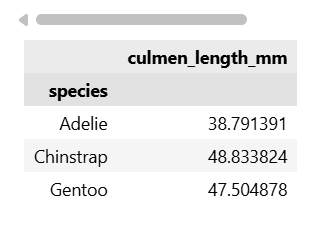
Nhận xét:

Adelie & Chinstrap: dữ liệu được thu thập rất đều, không thiên lệch giới tính.

Gentoo: Có sự chênh lệch nhẹ

4. Phân tích 2 biến sử dụng pivot\_table

Bảng 3: Độ dài mỏ trung bình theo loài

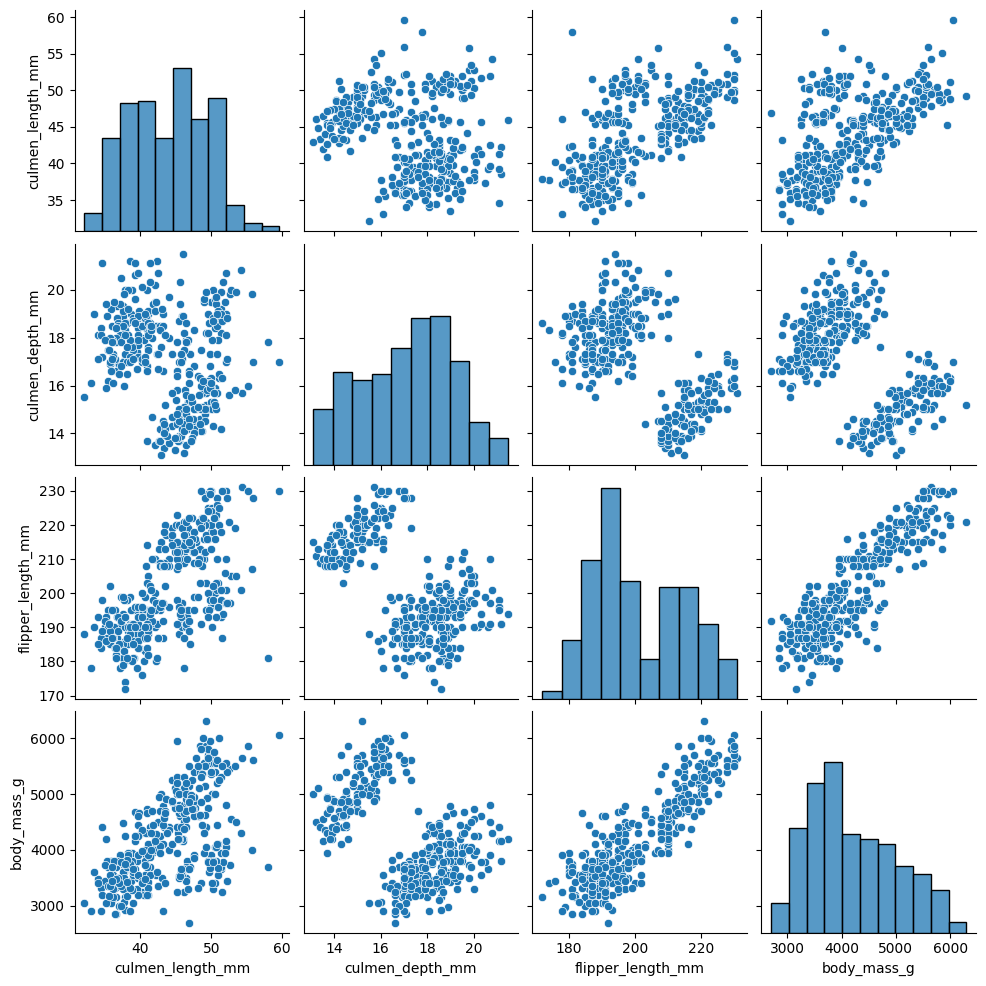


Nhận xét:

Gentoo có độ dài mỏ trung bình lớn nhất, tiếp theo là Chinstrap, và Adelie có mỏ ngắn nhất.

Điều này phù hợp với các biểu đồ bạn đã cung cấp trước đó, cho thấy Gentoo thường có kích thước cơ thể lớn hơn và mỏ dài hơn.

5. Phân tích 2 biến sử dụng pairplot



Hình 3: Biểu đồ pair-plot

Nhận xét:

Chim có mỏ dài thường có khối lượng cơ thể lớn hơn.

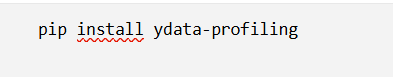
Chim có vây dài thì thường nặng hơn.

Chim có mỏ dài thì vây cũng thường dài.

**Bài toán 3:** Thực hiện các nhiệm vụ trong bài toán 3 để làm quen với việc sử dụng các công cụ hỗ trợ EDA tự động.

Nhiệm vụ 1: Sử dụng pandas profiling trên dữ liệu Customer Personality Analysis.

1. Cài đặt pandas\_profiling sau này đổi tên thành ydata\_profiling (xem thông tin chi tiết tại <https://pypi.org/project/pandas-profiling/3.1.0>)



Hình 4: Hướng dẫn tải thư viện

2. Sử dụng công cụ



3. Tiến hành EDA trên trang tập tin profile\_output.html