

Business Objective: Memahami Perilaku dan Populasi Penguin di Palmer Archipelago, Antartika

Dataset : [Palmer Archipelago \(Antarctica\) penguin data \(kaggle.com\)](https://www.kaggle.com/datasets/adelieandrews/palmer-archipelago-penguin-data)

- Deskripsi: Tujuan bisnisnya adalah untuk memahami lebih dalam tentang perilaku, populasi, dan karakteristik penguin yang berpopulasi di Palmer Archipelago, Antartika. Tujuan ini juga dapat mencakup upaya konservasi spesies penguin di wilayah ini.
- Analisis yang mungkin:
  - Analisis deskriptif tentang karakteristik fisik penguin, seperti jenis kelamin, berat badan, panjang paruh, dan lainnya.
  - Identifikasi faktor-faktor lingkungan yang memengaruhi dinamika populasi penguin.
  - Menilai kesehatan dan kondisi penguin di wilayah tersebut.
  - Rekomendasi untuk upaya konservasi spesies penguin di Palmer Archipelago.

EDA

```
Handling Missing Values

Cek jumlah missing value!

[ ] Input your code here
data.isnull().sum()

species      0
island       0
culmen_length_mm  2
culmen_depth_mm  2
flipper_length_mm  2
body_mass_g    2
sex          10
dtype: int64

Mengisi missing value pada sex dengan modus!

[ ] Input your code here
data.sex.fillna(data.sex.mode()[0], inplace=True)

Pada baris ke-336 kolom sex, terdapat data berisi NaN, drop baris data tersebut!

[ ] Input your code here
data = data.dropna(axis=0)
data = data.drop(336, axis=0)

Cek kembali apakah masih ada missing value!
```

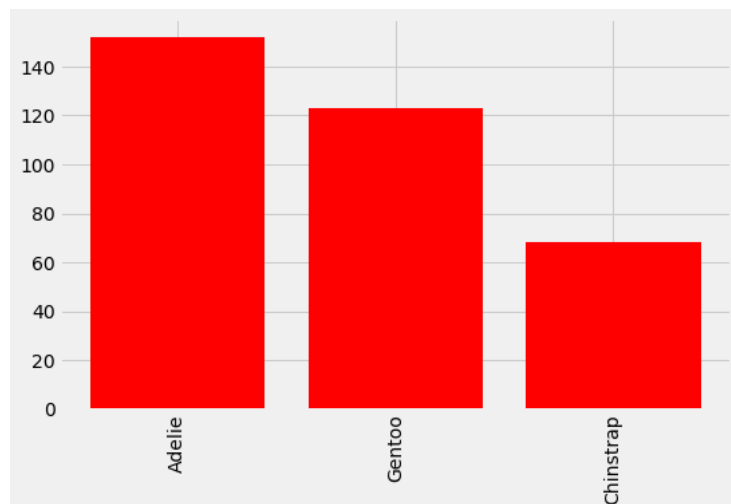
```
Cek kembali apakah masih ada missing value!

[ ] Input your code here
data.isnull().sum()

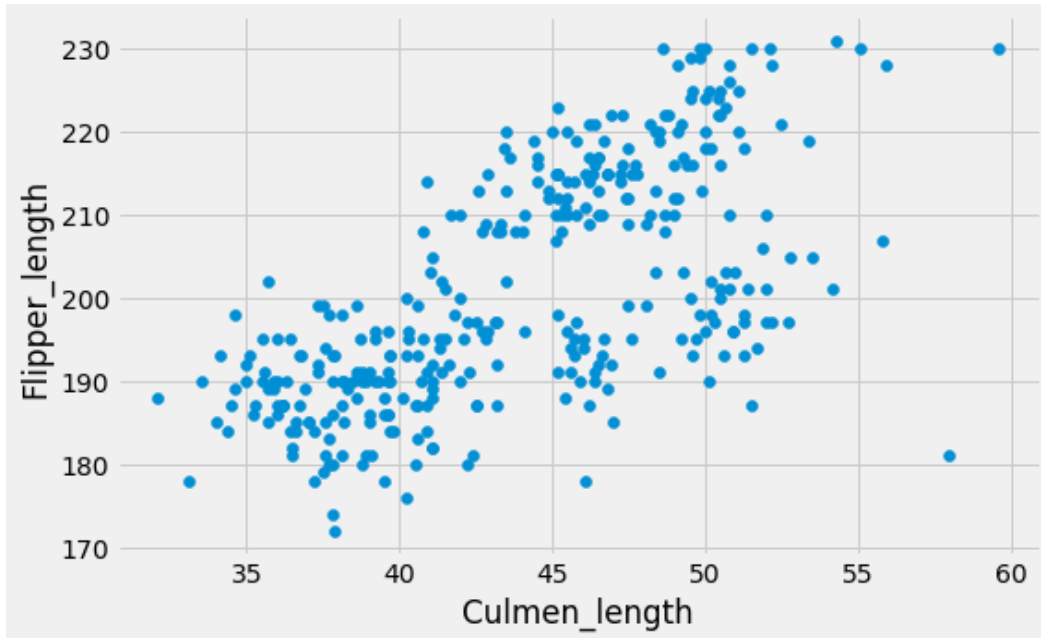
species      0
island       0
culmen_length_mm  2
culmen_depth_mm  2
flipper_length_mm  2
body_mass_g    2
sex          0
dtype: int64

[ ] data[['culmen_length_mm', 'culmen_depth_mm', 'flipper_length_mm']].dropna(inplace=True)
```

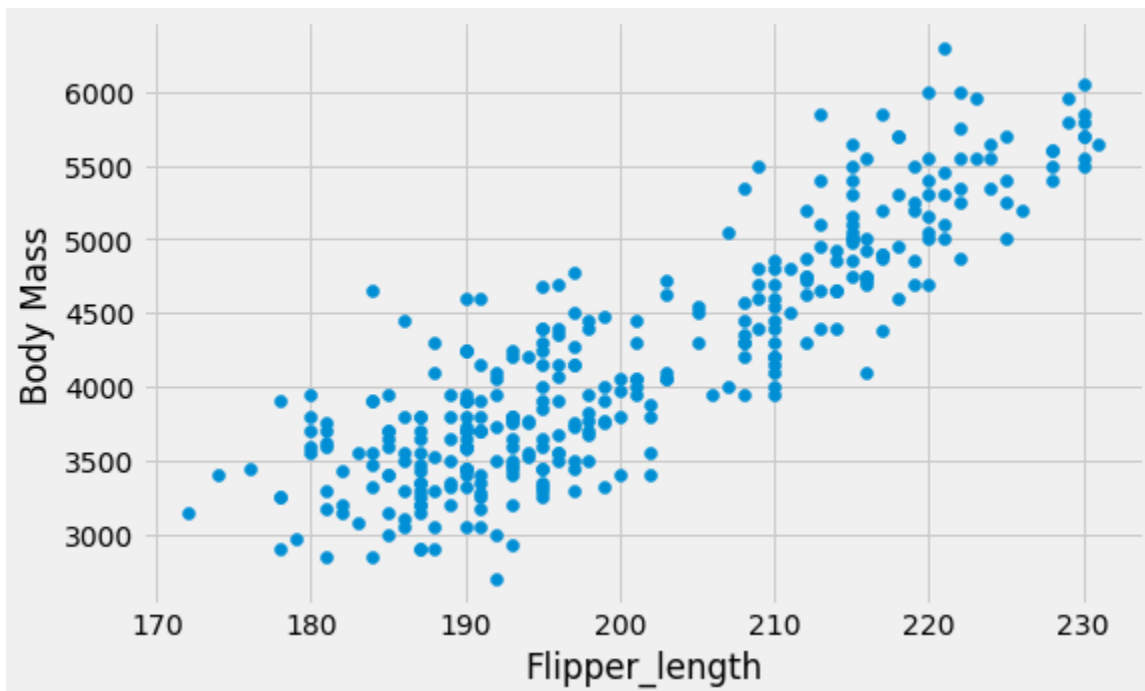
Visualisasi



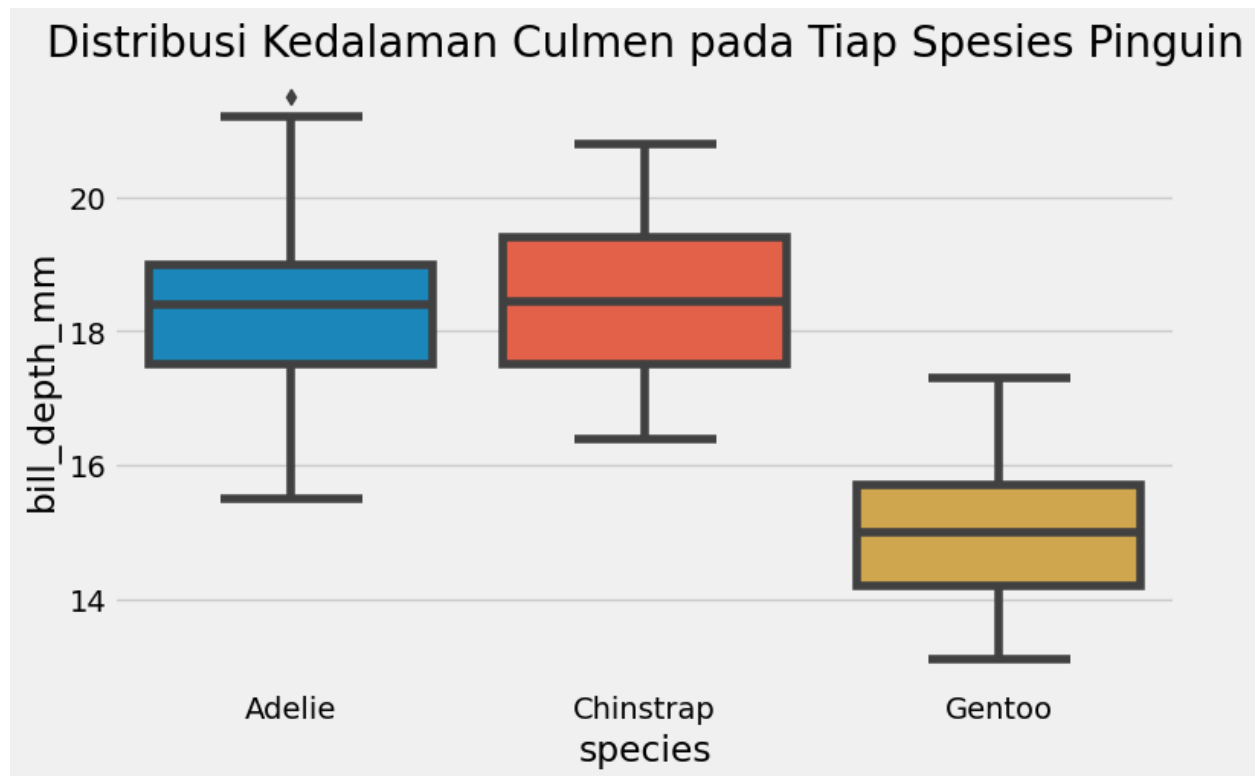
Spesies apa dengan jumlah penguin paling banyak? Adelie Penguin (*Pygoscelis adeliae*) dengan jumlah 152



Korelasi antara culmen\_length dan flipper\_length adalah lemah, karena titik titik tidak membentuk linier yang jelas atau tingkat pengaruhnya masih diragukan

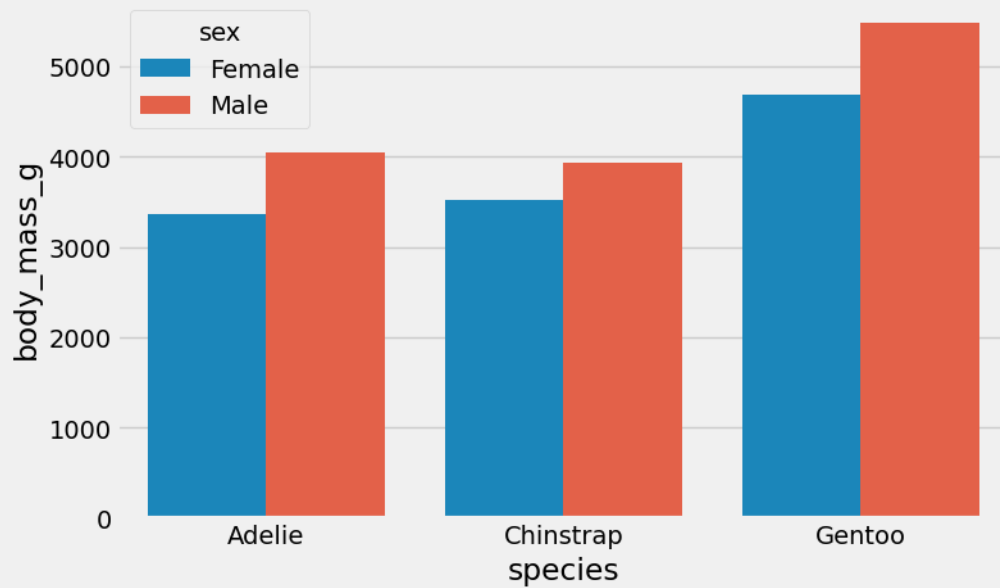


Korelasi antara flipper\_length dan body\_mass terlihat kuat, dikarenakan sebaran titik-titik mengelompok dalam bentuk linier yang jelas. flipper\_length mempengaruhi body\_mass secara langsung, jadi dapat disimpulkan bahwa setiap perubahan pada flipper\_length mempengaruhi perubahan pada body\_mass



- Terdapat tiga kotak (box) pada plot, masing-masing untuk spesies Adelie, Chinstrap, dan Gentoo.
- Garis horisontal di tengah setiap kotak adalah median (nilai tengah) dari distribusi data kedalaman culmen pada tiap spesies.
- Rentang data di dalam kotak (box) menunjukkan rentang interkuartil (IQR), yaitu jarak antara kuartil pertama (Q1) dan kuartil ketiga (Q3).
- Garis vertikal di atas dan di bawah kotak (box) disebut whisker. Whisker menunjukkan rentang data yang tidak dianggap sebagai outlier pada setiap spesies penguin.
- Tidak ada titik-titik di atas dan di bawah whisker, yang menandakan bahwa tidak ada outlier dalam data kedalaman culmen pada setiap spesies.
- Jarak antara kotak (box) dan whisker dapat memberikan informasi tentang variabilitas data di luar rentang interkuartil (IQR).
- Berdasarkan posisi median dan rentang data, terlihat bahwa spesies Gentoo memiliki nilai median yang lebih tinggi dibandingkan dengan spesies Adelie dan Chinstrap.
- Secara umum, spesies Gentoo cenderung memiliki kedalaman culmen yang lebih besar dibandingkan dengan spesies Adelie dan Chinstrap.-

Rata-rata Body Mass Pinguin berdasarkan Spesies dan Jenis Kelamin



[balqisalsya/analysis-penguins-dataset \(github.com\)](https://github.com/balqisalsya/analysis-penguins-dataset)