Bing

Voici une traduction possible de l'algorithme de clustering hiérarchique agglomératif en Python :

# Importer les bibliothèques nécessaires

import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import AgglomerativeClustering
from scipy.cluster import hierarchy

# Charger le jeu de données des pingouins

df = pd.read\_csv("penguins.csv")

# Sélectionner les colonnes qui seront utilisées pour le clustering

df = df[["bill\_length\_mm", "flipper\_length\_mm"]]

# Supprimer les lignes avec des données manquantes

df = df.dropna(axis=0)

# Effectuer le clustering hiérarchique agglomératif avec 3 clusters

clustering\_model = AgglomerativeClustering(n\_clusters=3, linkage="ward")
clustering\_model.fit(df)
labels = clustering\_model.labels\_

# Afficher le dendrogramme des sous-clusters

clusters = hierarchy.linkage(df, method="ward")
plt.figure(figsize=(8, 6))
dendrogram = hierarchy.dendrogram(clusters)
plt.title("Dendrogramme des sous-clusters")
plt.show()

# Afficher les données avant et après le clustering

fig, axes = plt.subplots(nrows=1, ncols=2, figsize=(15, 5))
axes[0].scatter(df["bill\_length\_mm"], df["flipper\_length\_mm"])
axes[0].set\_title("Sans clustering")
axes[0].set\_xlabel("Longueur du bec (mm)")
axes[0].set\_ylabel("Longueur des palmes (mm)")
axes[1].scatter(df["bill\_length\_mm"], df["flipper\_length\_mm"], c=labels)
axes[1].set\_title("Avec clustering")
axes[1].set\_xlabel("Longueur du bec (mm)")
axes[1].set\_ylabel("Longueur des palmes (mm)")
plt.show()