1. **Khái niệm và mục tiêu của K-Means.**

Một trong những thuật toán phổ biến được sử dụng để phân cụm là thuật toán K-Means.

Vì không có nhãn trong dữ liệu chưa được gắn nhãn, do đó chúng tôi quan tâm đến việc chúng tôi có thể tìm thấy các mẫu trong dữ liệu chưa được gắn nhãn đó. Kỹ thuật tìm kiếm các mẫu trong dữ liệu không được gắn nhãn này được gọi là phân cụm. Mục đích chính của phân cụm là phân tích các nhóm cổng có các đặc điểm giống nhau và gán chúng thành các nhóm (thường được gọi là các cụm).

K-Means clustering là một loại học tập không giám sát:

+ Được sử dụng khi bạn có dữ liệu chưa được gắn nhãn.

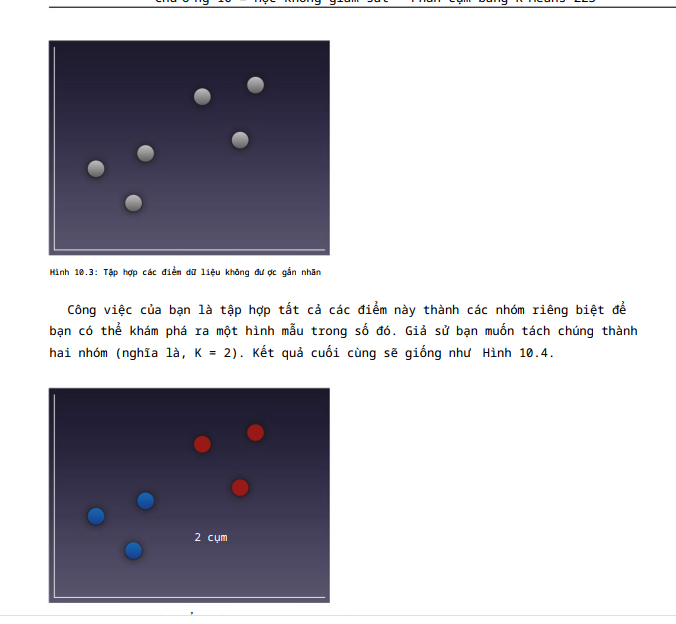
+ Mục tiêu là tìm các nhóm trong dữ liệu, với số lượng các nhóm được đại diện bới K.

Mục tiêu của phân cụm K-Means là đạt được những điều sau

+ K centroid đại diện cho trung tâm của các cụm

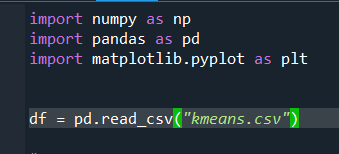
+ Nhãn cho dữ liệu đào tạo

1. **Ví dụ**

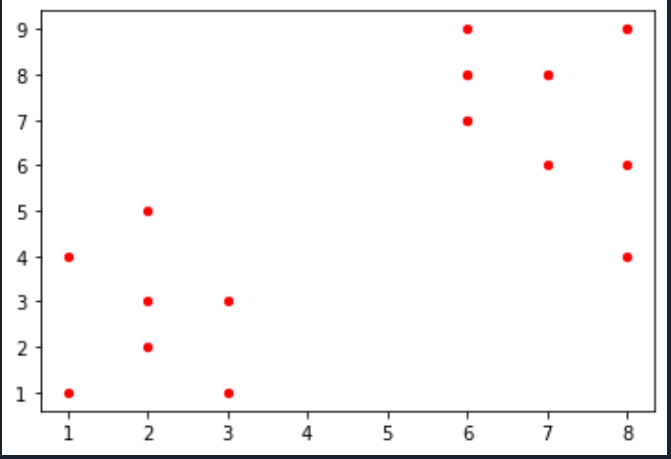
****

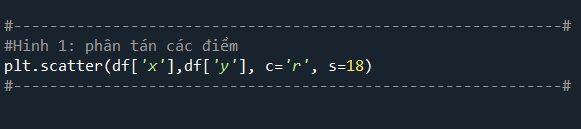
1. **Code**

**Khai báo thư viện**

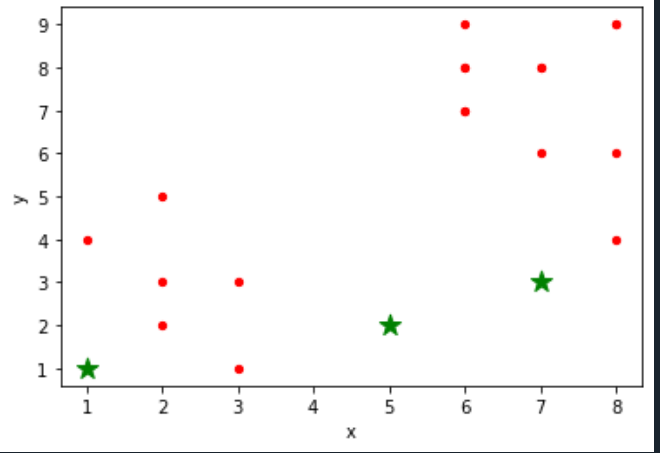
****

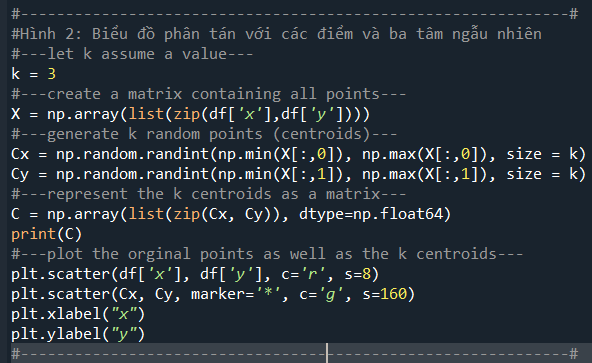
**Biểu đồ phân tán các điểm**

****

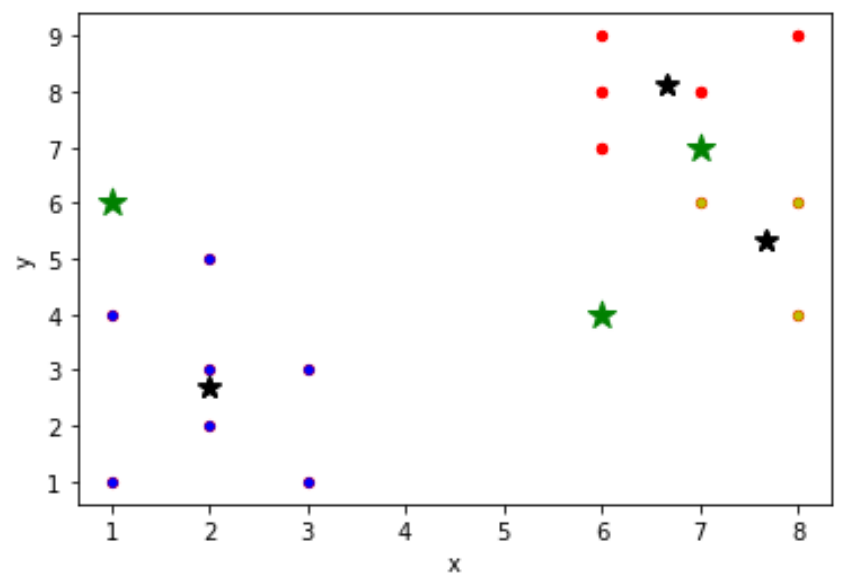
****

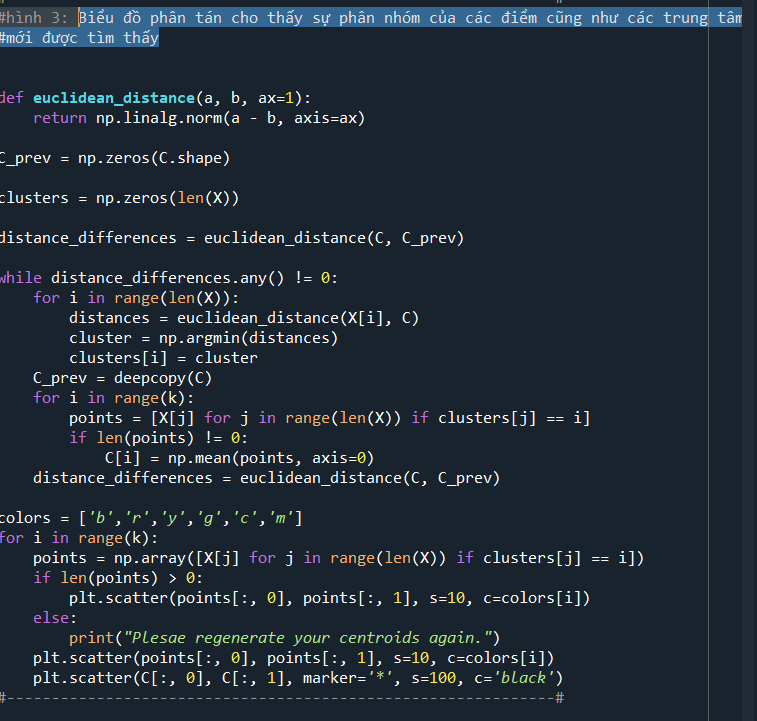
**Biểu đồ phân tán với các điểm và ba tâm ngẫu nhiên (k=3)**

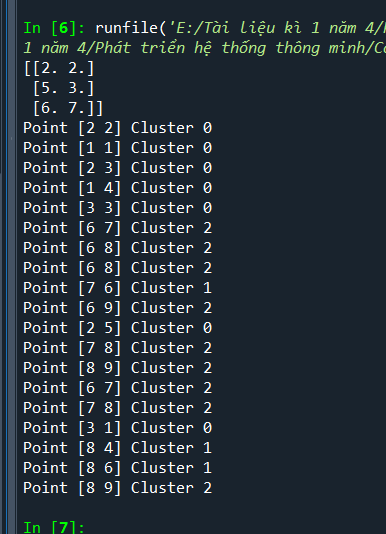
****

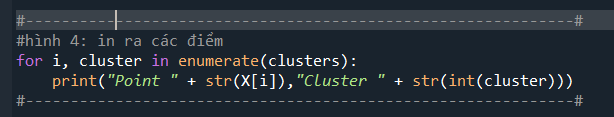
****

**Biểu đồ phân tán cho thấy sự phân nhóm của các điểm cũng như các trung tâm mới được tìm thấy**

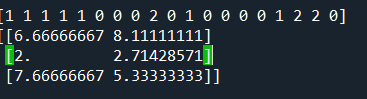
****

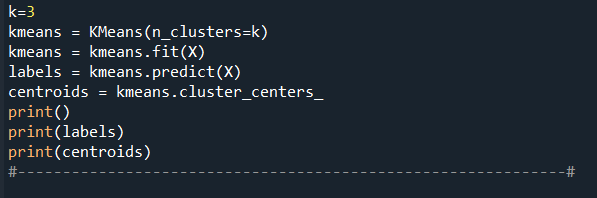
****

**in ra các điểm và giá trị C**

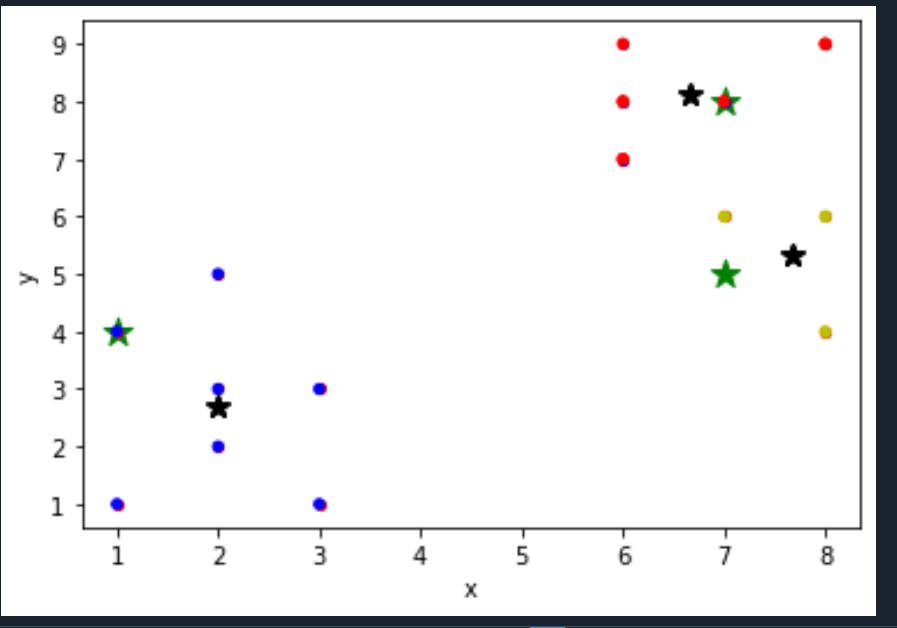
****

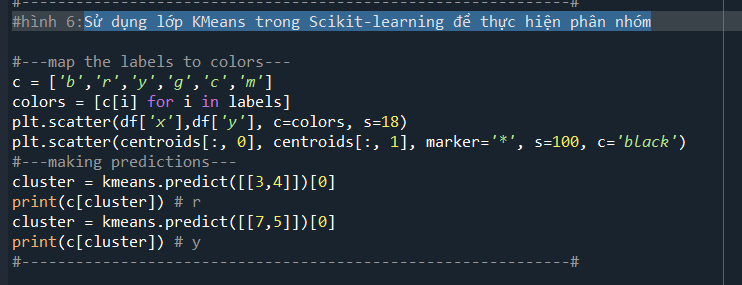
**Dữ liệu**

****

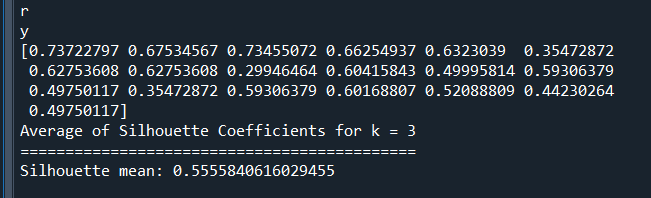
****

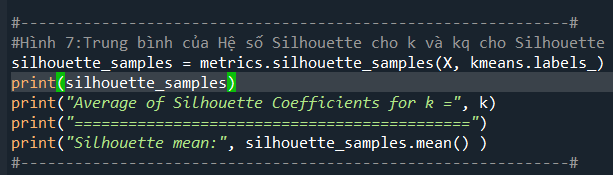
**Sử dụng lớp KMeans trong Scikit-learning để thực hiện phân nhóm**

****

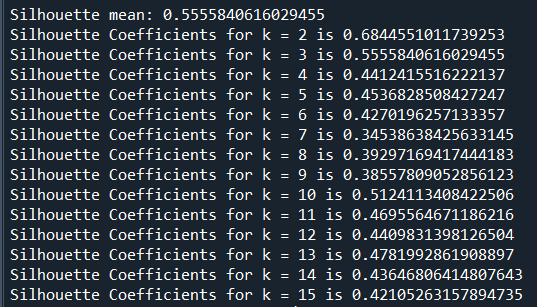
****

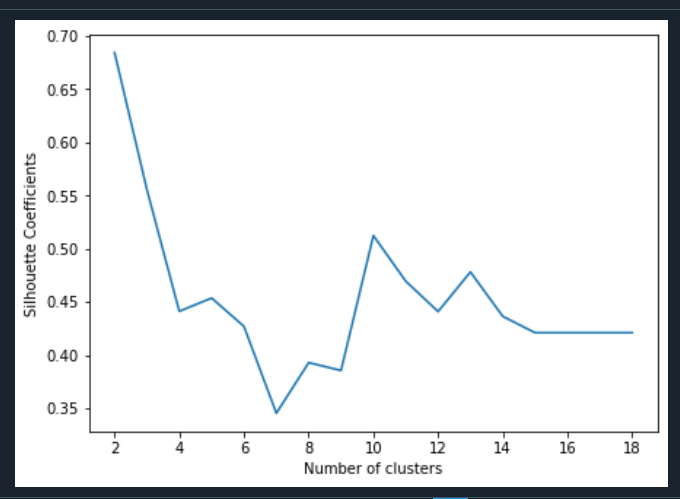
**#Hình 7:Trung bình của Hệ số Silhouette cho k và kq cho Silhouette**

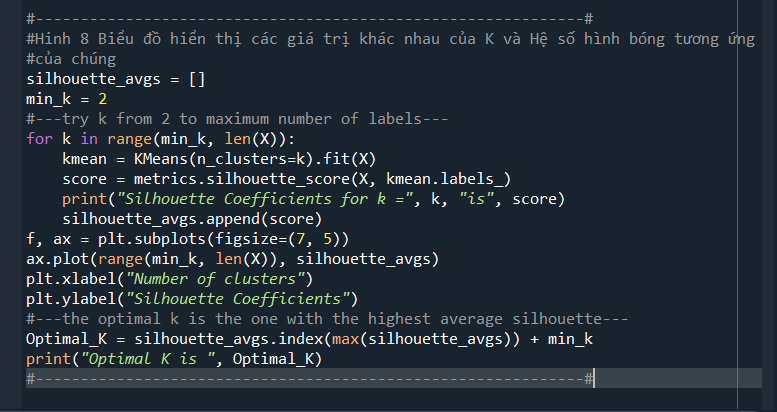
****

****

**#Hinh 8 Biểu đồ hiển thị các giá trị khác nhau của K và Hệ số hình bóng tương ứng của chúng.**

****

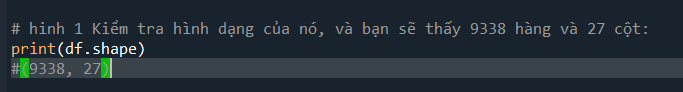
****

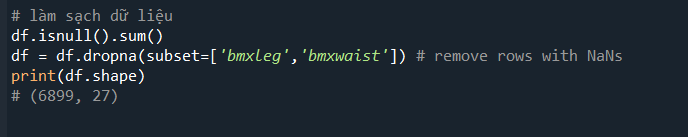
****

Giả sử bạn là một nhà thiết kế quần áo và bạn được giao nhiệm vụ thiết kế một loạt quần short Bermuda mới. Một trong những vấn đề thiết kế là bạn cần phải đưa ra một loạt các kích thước sao cho phù hợp với hầu hết mọi người. Về cơ bản, bạn cần có một loạt quy mô của những người khác nhau:

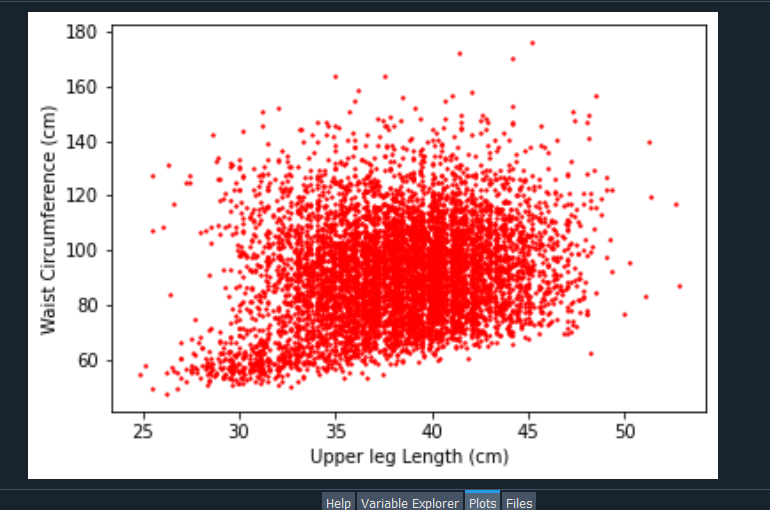
+ chu vi vòng eo

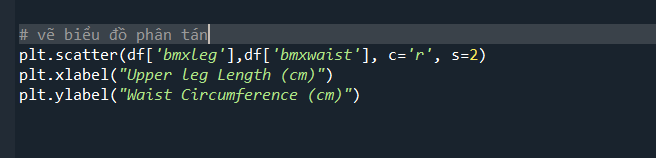
+ chiều dài chân trên

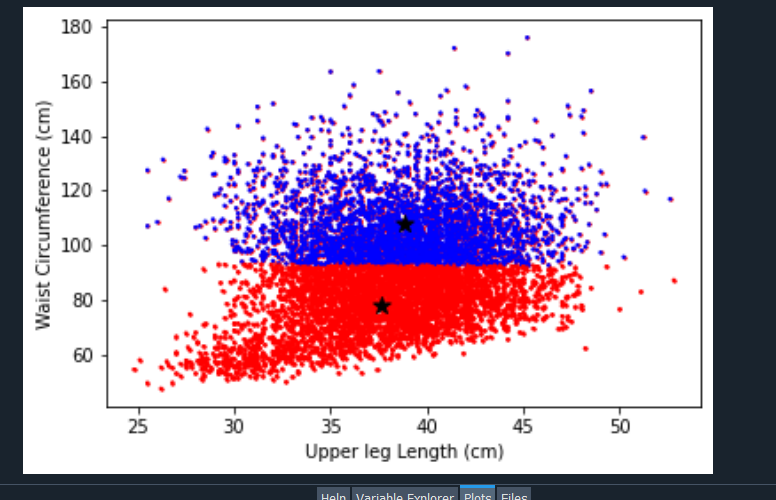
****

****

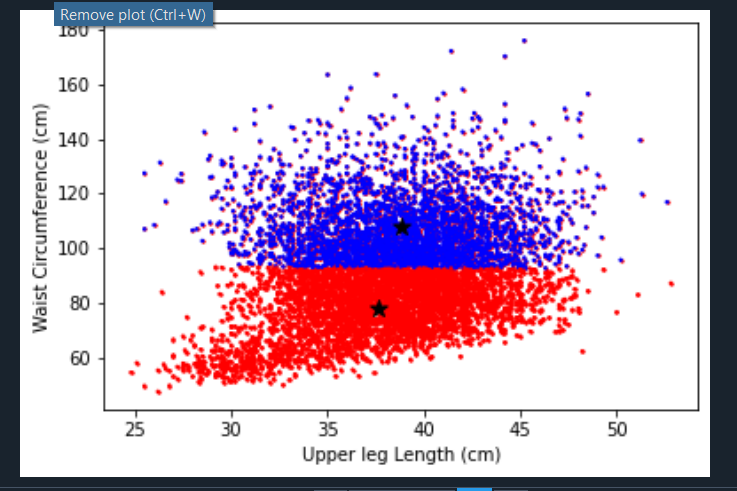
Biểu đồ phân tán cho thấy sự phân bố của chu vi vòng eo và chiều dài chân trên

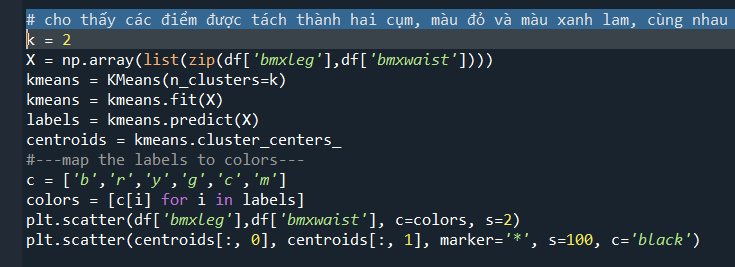
****

****

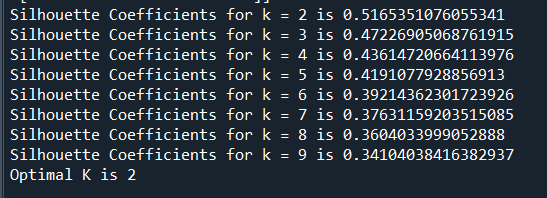
****

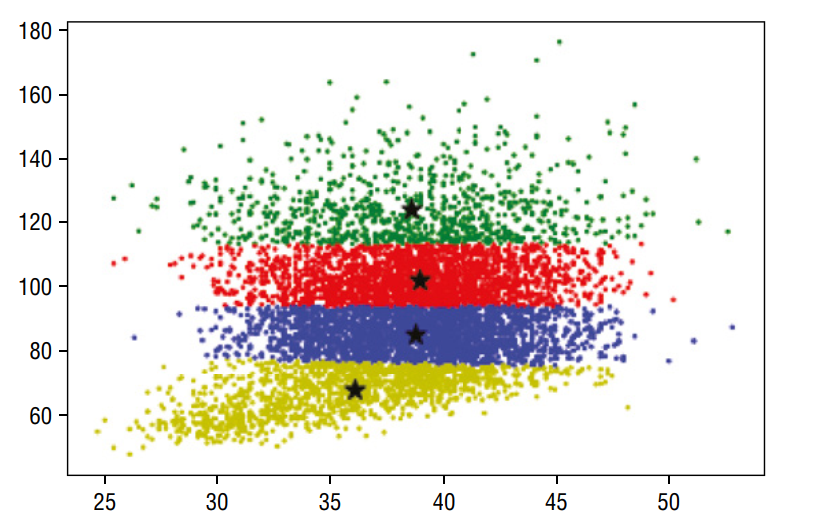
**# cho thấy các điểm được tách thành hai cụm, màu đỏ và màu xanh lam, cùng nhau với hai trung tâm.**

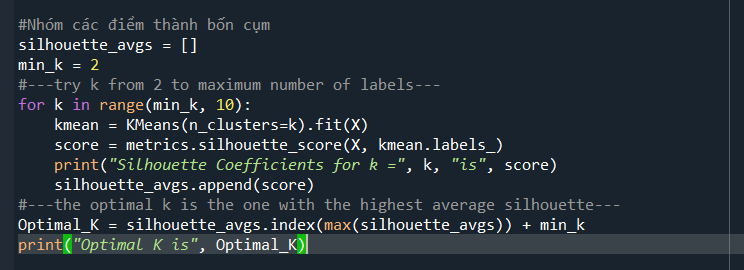
****

****

**#Nhóm các điểm thành bốn cụm**

****

****

****