

I.

Bài 1 em vào thẳng trang <https://fbref.com/en/comps/9/2023-2024/playingtime/2023-2024-Premier-League-Stats> để cào dữ liệu các cầu thủ chứ không qua từ CLB. Bảng "Player Playing Time" được tải động bằng JavaScript sau khi trang HTML ban đầu được tải về nên em phải dùng selenium mới cào được dữ liệu

II.

Theo bạn đội nào có phong độ tốt nhất giải ngoại Hạng Anh mùa 2023-2024: Câu này em xem đội nào có nhiều chỉ số cao nhất là đội có phong độ tốt nhất

III.

Theo bạn thì nên phân loại cầu thủ thành bao nhiêu nhóm?

Dựa trên biểu đồ Elbow của phương pháp xác định số cụm tối ưu, ta sẽ chọn số cụm tại vị trí mà đường cong bắt đầu "uốn cong" và trở nên ít dốc hơn, tức là điểm mà giảm thiểu của WCSS (Within-Cluster Sum of Squares) không còn quá lớn khi tăng số cụm. Trong trường hợp này, vị trí "uốn cong" của đường cong xuất hiện ở khoảng $k = 3$ hoặc $k = 4$. Vậy nên ta sẽ chọn chia thành 3 nhóm.

