- Phần xử lí dữ liệu: sử dụng Python

- Phần biểu diễn dữ liệu: sử dụng PowerBI

\_\_ PYTHON\_\_

1. Bảng top 250

- Sắp xếp lại dữ liệu qua hàm sort

- Đọc các bảng và xem qua trạng thái các bảng, cột để tìm các giá trị lỗi, các giá trị null

- Kiểm tra xem có giá trị bị trùng hay lặp không

- Chỉnh sửa lại các giá trị viết tắt ở cột name bảng top 250

- Điền các giá trị lỗi ở cột age bảng top 250 bằng hàm mean()

- Loại bỏ các cột không cần thiết như cột Market\_value vì dữ liệu đã cũ

- Kiểm tra lại số liệu và export thành file top250clean

2. Bảng trophies

- Đọc dữ liệu và xem qua trạng thái các bảng, cột để tìm giá trị lỗi, các giá trị null

- Thêm các giá trị còn thiếu bằng cách sử dụng số trophies trước đấy ( Ví dụ: 2001 : Arsenal có 35 trophies, 2002: Arsenal null -> 2002: Arsenal có 35 trophies

- Chỉnh sửa lại tên đội sao cho thống nhất với bảng top250

- Điều chỉnh lại bố cục bảng sao cho phù hợp với việc sử dụng PowerBI sau này

- Kiểm tra lại số liệu và export thành file trophiesclean

\_\_POWERBI\_

- Kiểm tra lại dữ liệu và nối relationship (duplicate bảng trophiesclean và nối dựa trên season và football teams)

1. Page 1

- Sử dụng table, chart và line graph để biểu diễn nội dung tổng quát (sử dụng 1 vài measure như count(), sum())

- Có thể xem cụ thể từng positions hoặc cầu thủ bằng cách click vào giá trị trong chart hoặc table để xem tổng quát

2. Page 2

- Xem cụ thể Top N Player Transfer (filter bằng độ tuổi, vị trí, mùa giải), sử dụng các measure như sum(), TopN(), SelectedValue(),…

3. Page 3

- Xem thành tích của 1 vài đội bóng và sự liên quan đến các kì chuyển nhượng