Tên Sinh Viên : Trần Đoàn Ngọc Duy

Mã số sinh viên: 20E1020014

Sinh viên Khóa 1 khoa Kỹ Thuật và Công Nghệ

Đồ Án môn: Phân Tích dữ liệu với R

Một số khái niệm:

Bàn thắng kỳ vọng(xG): là tổng các tỉ lệ để có được bàn thắng, tùy vào vị trí, cầu thủ, thời tiết,… để tính

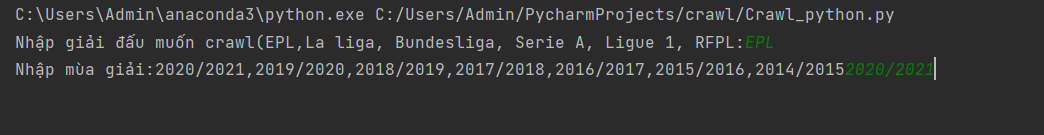
Bàn thua kỳ vọng(xGA): Ngược lại với bàn thắng kỳ vọng

Những khái niệm kỳ vọng khác được tính dựa theo bàn thắng và bàn thua kỳ vọng y như bàn thắng bàn thua.

# Crawl Dữ liệu:

Em dùng selenium để tự động vào trang web: <https://understat.com/> để lấy số liệu về kết quả những trận đấu và bàn thắng, bàn thua kỳ vọng của từng trận đấu

Bài code của em : <https://github.com/d2525456d/CrawlData/blob/main/crawl/Crawl_python.py>

Chạy code 

Nhập vào giải đấu muốn crawl và mùa giải cần crawl.

(em lấy EPL mùa giải 2020/2021 để phân tích)

Chạy xong ta được 1 file csv

<https://github.com/d2525456d/CrawlData/blob/main/crawl/File_craw4446.csv>

# Xử lý số liệu để crawl

Code của em:

<https://github.com/d2525456d/CrawlData/blob/main/crawl/Xulysolieu.ipynb>

Đầu tiên em đưa file csv ở trên vào jupyter để xử lý số liệu. Dữ liệu em xử lý nhằm có thể phân tích, làm thống kê mô tả, có ý nghĩa về mặt thống kê.

Dữ liệu được đưa về 1 cái Bảng xếp hạng gồm các clb, điểm, bàn thắng, bàn thắng kỳ vọng,… để có thể thống kê. Rồi lưu về thành 1 file csv.

Chi tiết bài xử lý ở bài code.

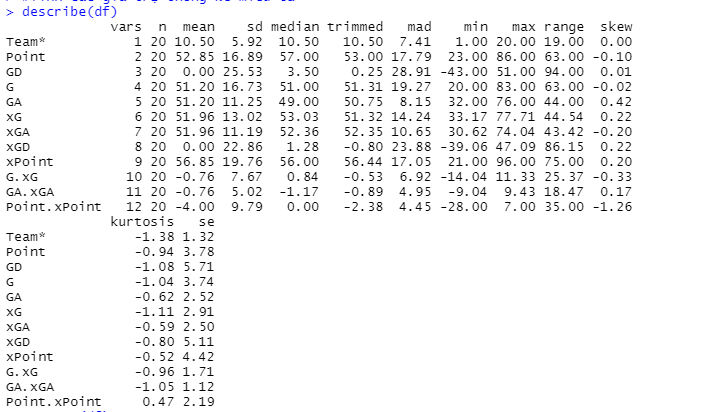
# Phân tích thống kê mô tả

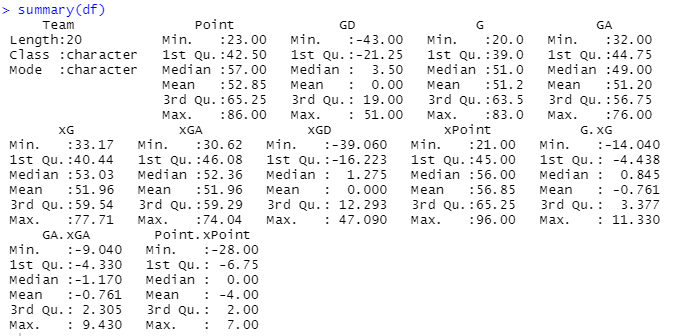
Code của em :

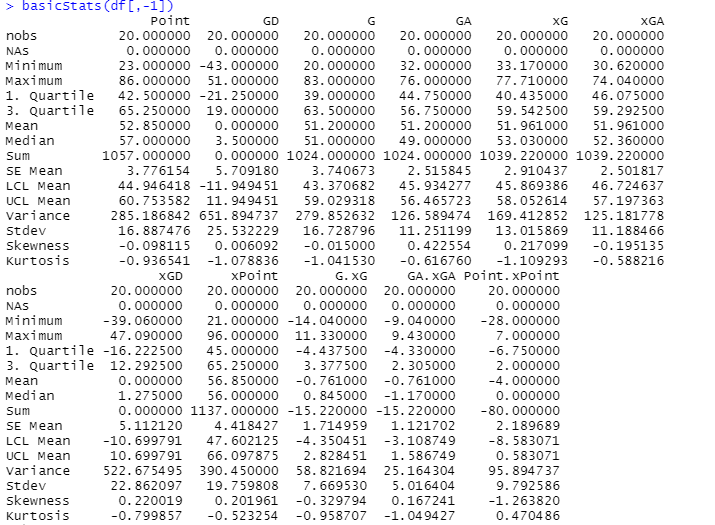
<https://github.com/d2525456d/CrawlData/blob/main/crawl/Thongkemota.R>

Đầu tiên set về thư mục để làm việc, sau đó em đưa file csv đã xử lý vào.

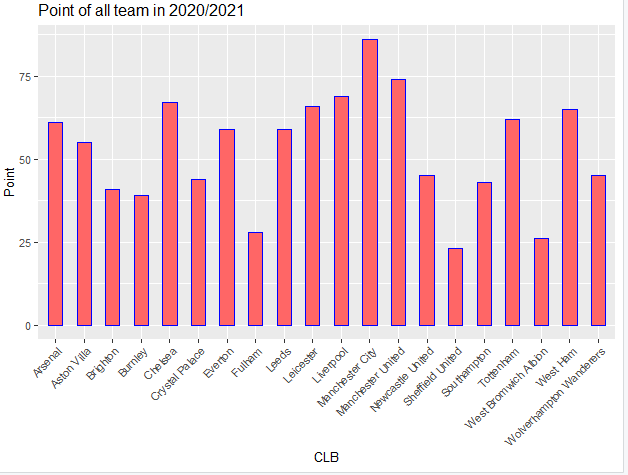
Dùng những hàm describe, summary, basicStats





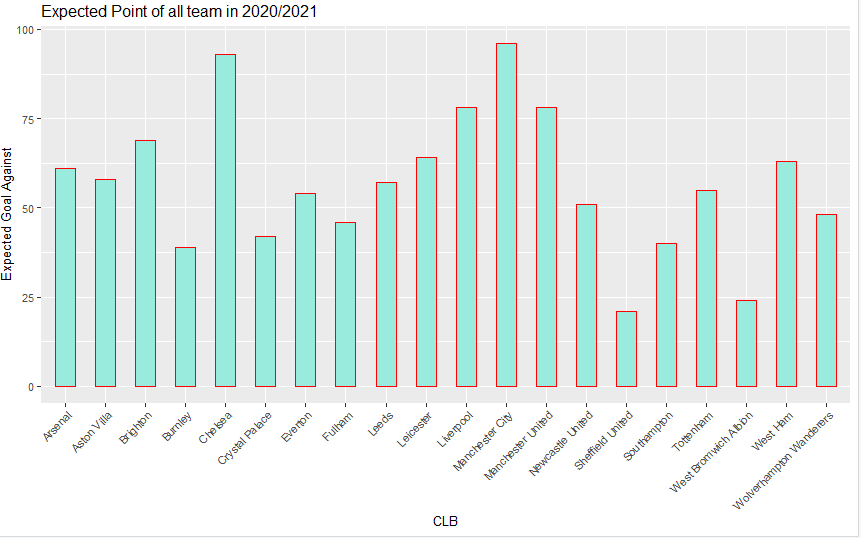


Vẽ biểu đồ :



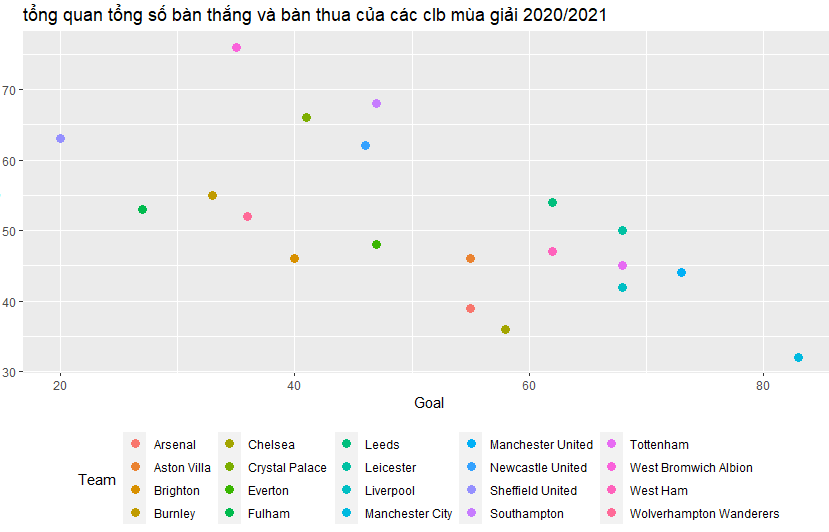
Nhận xét:

* Manchester city là đội bóng vô địch.
* Sheffield United, West Bromwich Albion, FullHam là 3 đội bóng bị xuống hạng.
* Theo bảng xếp hạng ( không kể vô địch cup FA, C1, C2, C3) thì Manchester City, Manchester United, Liverpool, Chelsea là 4 đội bóng tham dự cup C1, Leicester City và West Ham tham dự C2 và Tottenham tham dự C3.



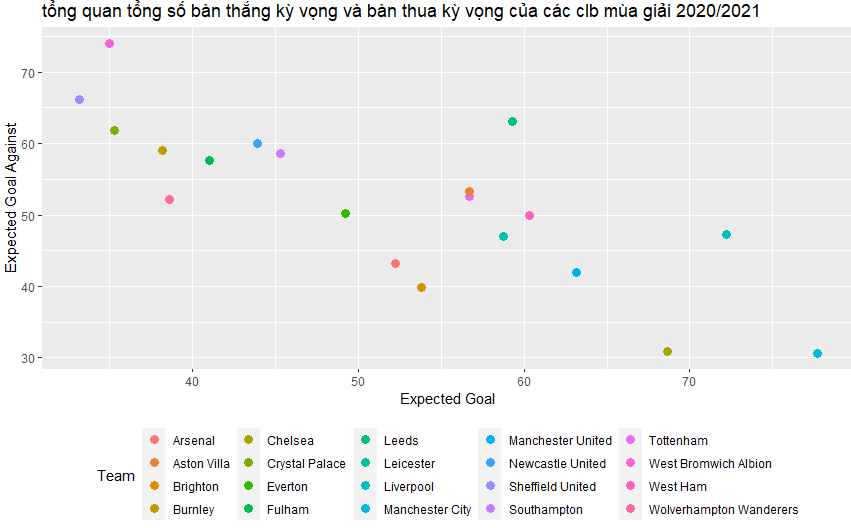
Nhận xét :

* Nếu tính theo điểm kỳ vọng thì đội vô địch vẫn là Manchester City.
* Có một số xáo trộn so với điểm thực.
* 4 đội tham dự C1 vẫn là Manchester City, Chelsea, Liverpool, Manchester United.
* 2 đội tham dự C2 theo điểm kỳ vọng là Brighton và Leicester.
* Đội tham dự C3 là West Ham.
* 3 đội bóng xuống hạng nếu tính theo điểm kỳ vọng là Sheffield United, West Bromwich Albion và Burnley.
* Có khá nhiều thay đổi so với điểm thực.



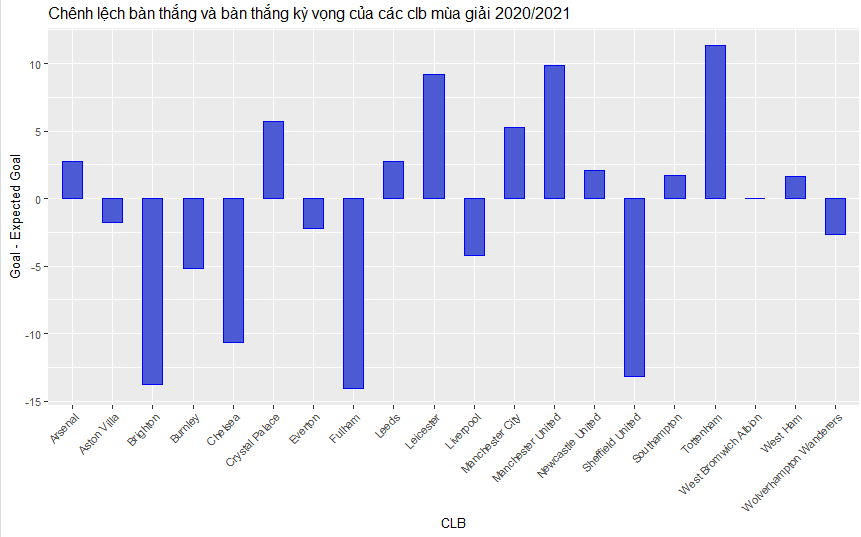
Nhận xét :

* Manchester city đứng đầu về số bàn thắng và có ít bàn thua nhất.
* Sheffield United có số bàn thắng rất ít nhưng số bàn thua khá cao.
* West Bromwich Albion có số bàn thắng không quá ít nhưng nhận rất nhiều bàn thua dẫn đến phải rớt hạng
* Có sự phân hóa khá rõ ràng giữa những đội có hiệu số tốt.



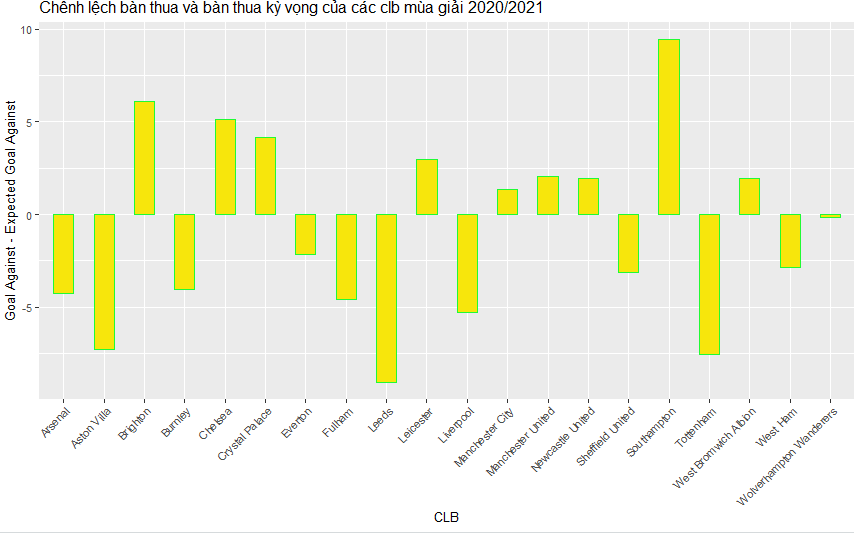
Nhận xét :

* Tính theo kỳ vọng thì Manchester City cũng là đội bóng có được ít bàn thua và nhiều bàn thắng nhất.
* West Bromwich Albion là đội bóng có bàn thua kỳ vọng rất nhiều, bàn thắng kỳ vọng cũng rất ít ( nhiều hơn mỗi Sheffield United).
* Chelsea và Brighton cho thấy sự thay đổi rất nhiều.
* Thay đổi rất nhiều so với kết quả thực.



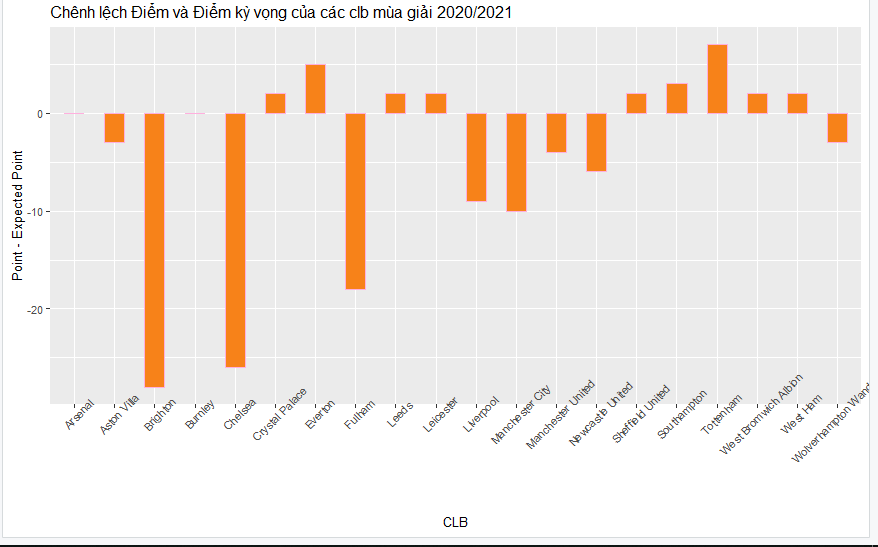
Nhận xét:

* Tottenham là đội bóng ghi được nhiều bàn thắng hơn so với bàn thắng kỳ vọng nhiều nhất. Chứng tỏ tiền đạo có khả năng săn bàn rất tốt.
* Tiếp sau đó là Manchester United và Leicester.
* Fullham, Brighton, Sheffield United là 3 đội bóng có ít bàn thắng hơn so với bàn thắng kỳ vọng nhiều nhất, chứng tỏ rằng 3 đội bóng này có hàng công rất tệ khi bỏ lỡ rất nhiều cơ hội dẫn đến vị trí khá thấp trong bảng xếp hạng. Cần nên tăng cường năng lực ở các vị trí tấn công.
* West Bromwich Albion là đội bóng gần như không có chênh lệch giữa bàn thắng và bàn thắng kỳ vọng.



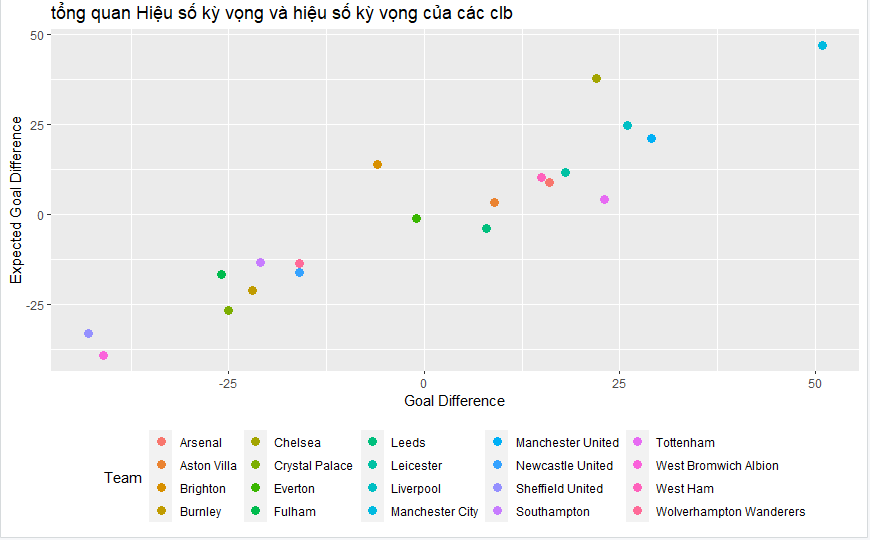
Nhận xét:

* Southampton là đội bóng có nhiều bàn thua so với bàn thua kỳ vọng nhiều nhất. Điều này chứng tỏ hàng thủ và đặc biệt là thủ môn thi đấu khá tệ, nên củng cố hơn hàng phòng ngự, đặt biệt là thủ môn
* Leeds, Aston Villa, Tottenham là 3 đội bóng có ít bàn thua so với bàn thua kỳ vọng nhiều nhất, chứng tỏ hàng thủ và thủ môn làm việc rất hiệu quả
* Wolverhampton Wanderers hầu như không có sự chênh lệch giữa bàn thua và bàn thua kỳ vọng



Nhận xét:

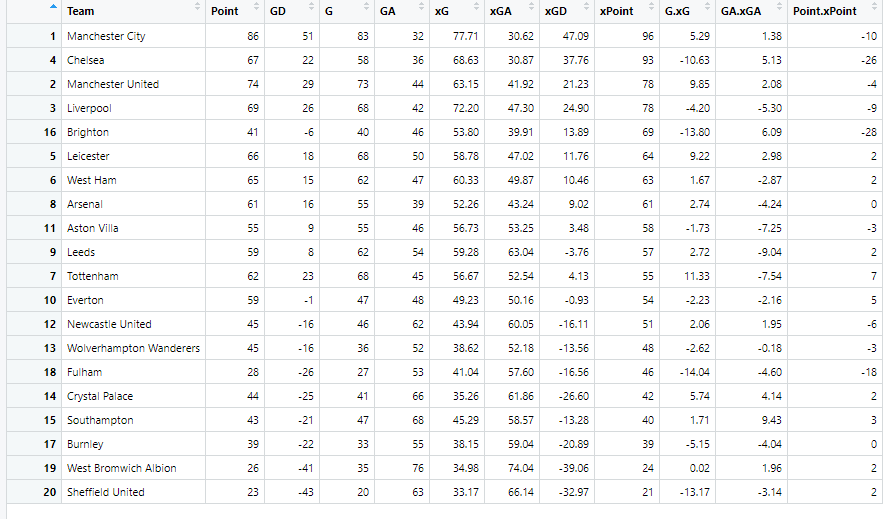
* Southamton, Eveton là những đội bóng có nhiều điểm hơn so với điểm kỳ vọng nhiều nhất, làm cải thiện thứ hạng của những đội bóng này rất nhiều.
* Brighton, Chelsea bị mất điểm rất nhiều khi hàng tấn công không tận dụng được nhiều cơ hội hoặc là hàng thủ không làm việc tốt dẫn đến thứ hạng của những đội này bị tụt khá nhiều.
* Arsenal và Burnley là 2 đội bóng hầu như không có chênh lệch điểm số so với điểm số kỳ vọng.



Nhận xét:

* West Bromwich albion và Shelffield United là 2 đội bóng có cả hiệu số và hiệu số kỳ vọng vô cùng thấp.
* Manchester City là một sự khác biệt khi hiệu số-hiệu số kỳ vọng ở rất cao.
* Những chấm của các đội đi qua hoặc sai số ít với đường thẳng y = x

Bảng xếp hạng theo kỳ vọng:



Nhận xét:

* Ta có thể thấy được nếu như theo kỳ vọng thì Brighton tăng lên tận 11 bậc.
* Manchester City vẫn là nhà vô địch
* Burnley là đội bóng xuống hạng thay vì Fulham
* Trừ một số trường hợp như Brighton, Aston Villa, Tottenham thì tương quan vị trí trên bảng xếp hạng kỳ vọng so với bảng xếp hạng ban đầu khá tương đồng.

--------------------------------------------Hết---------------------------------------------