NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG NHỮNG V

Không thể truy vấn người dùng% 1

i-I-I- 6-30-30

* Trong bài thí nghiệm này, tôi sử dụng các bộ dữ liệu Best Basketballs Stats để nghiên cứu và phân tích.

1. Bộ phát tích tích tích tích tích tích tích tích tích tích tích tích học

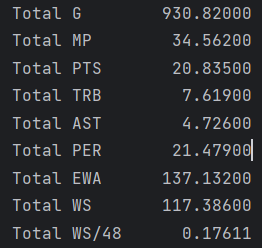
* Trò chơi trò chơi bóng bóng đến chơi bài tốt nhất
* Đối với mỗi bộ dữ liệu, xác định các tính năng nào là không liên tục hoặc không liên tục? đánh giá hay định lượng? ut? ut hoặc mt? ut hoặc ut? giải thích.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Feature** | **Discrete/Continuous** | **Quantitative/Qualitative** | **Numerical/Categorical** | **Explanation** |
| **Name** | N/A | Qualitative | Categorical | Names are labels, not numerical values. |
| **Pos** | N/A | Qualitative | Categorical | Positions (e.g., F, C, PG) are categories, not numerical values. |
| **Drafted** | Discrete | Quantitative | Numerical | The year drafted is a discrete numerical value (whole numbers). |
| **Retired** | N/A | Qualitative | Categorical | Binary variable (Y/N), not numerical. |
| **Pick** | Discrete | Quantitative | Numerical | Draft pick number is a discrete numerical value (whole numbers). |
| **Team** | N/A | Qualitative | Categorical | Team names are categories, not numerical values. |
| **Total G** | Discrete | Quantitative | Numerical | Total games played is a discrete numerical value (whole numbers). |
| **Total MP** | Continuous | Quantitative | Numerical | Total minutes played is a continuous numerical value (can be fractional). |
| **Total PTS** | Continuous | Quantitative | Numerical | Total points scored is a continuous numerical value (can be fractional). |
| **Total TRB** | Continuous | Quantitative | Numerical | Total rebounds is a continuous numerical value (can be fractional). |
| **Total AST** | Continuous | Quantitative | Numerical | Total assists is a continuous numerical value (can be fractional). |
| **Total PER** | Continuous | Quantitative | Numerical | Player Efficiency Rating is a continuous numerical value (can be fractional). |
| **Total EWA** | Continuous | Quantitative | Numerical | Estimated Wins Added is a continuous numerical value (can be fractional). |
| **Total WS** | Continuous | Quantitative | Numerical | Win Shares is a continuous numerical value (can be fractional). |
| **Total WS/48** | Continuous | Quantitative | Numerical | Win Shares per 48 minutes is a continuous numerical value (can be fractional). |

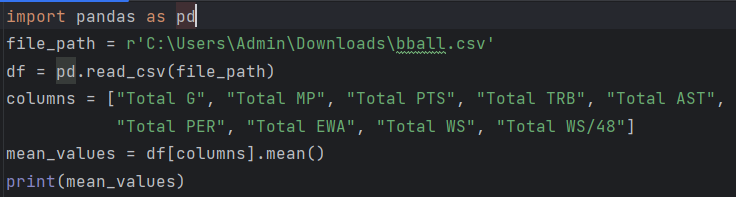
* Các vấn đề nằm bên trong dữ liệu là gì? Ví dụ: thiếu dữ liệu, dữ liệu sai? bạn có cần phải chuẩn bị dữ liệu?

Bộ dữ liệu xuất hiện đầy đủ với giá trị không thiếu hay sai. Tuy nhiên, việc xác minh độ chính xác đối với một nguồn đáng tin cậy được khuyên. Nếu thiếu dữ liệu hiện tại, chúng ta có thể gỡ bỏ hoặc gỡ bỏ các kí tự bị ảnh hưởng hoặc không tiền tố bằng cách dùng các tính năng và chế độ tiềm năng tiềm năng, hoặc một nơi cho những người dùng.

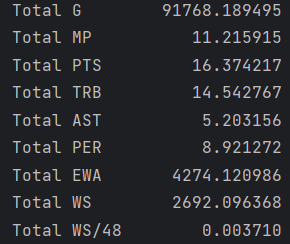
* Có dấu hiệu nào trong bộ dữ liệu không?
* Số: Bộ dữ liệu không có nhãn cụ thể ngụ ý biến mục tiêu cho dự đoán. Tuy nhiên, nếu bộ dữ liệu này được dùng cho một nhiệm vụ có tính năng dự đoán, một nhãn có thể được tạo ra dựa trên một tính năng đặc trưng.
* Ví dụ, tính năng tính năng, tính chất tương tác, khả năng của các dữ liệu đã chọn. Làm thế nào bạn có thể tính toán các biện pháp này để tính toán các tính năng đặc biệt (nếu có)?
* Giá trị trung bình của các tập hợp dữ liệu là:



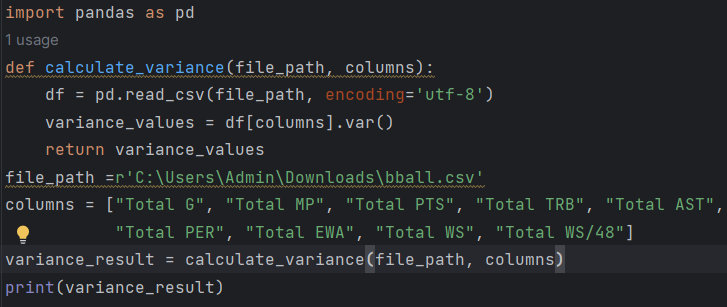
* & l#224;m l#224; m#7897;t c#225;i n#224;y.



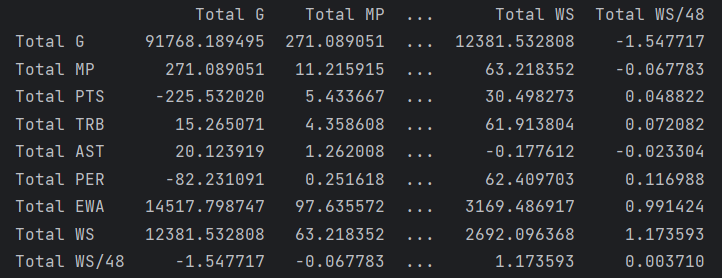
* Thành tích của các tập dữ liệu là:



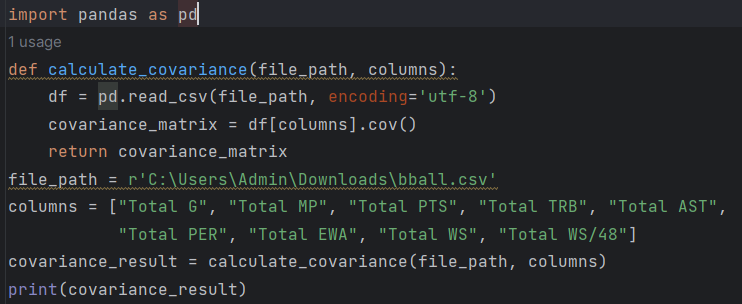
* Tính toán để tính toán các ký tự:



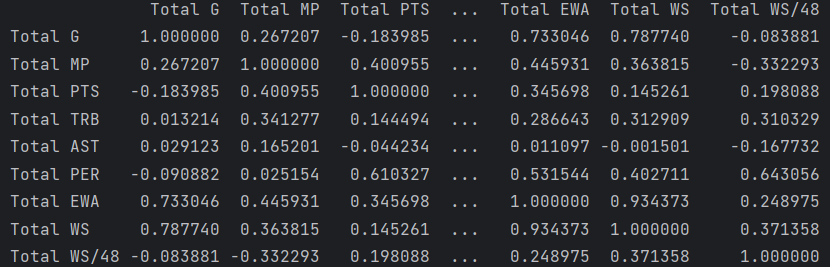
* Sự đồng thuận của các tập hợp dữ liệu là:



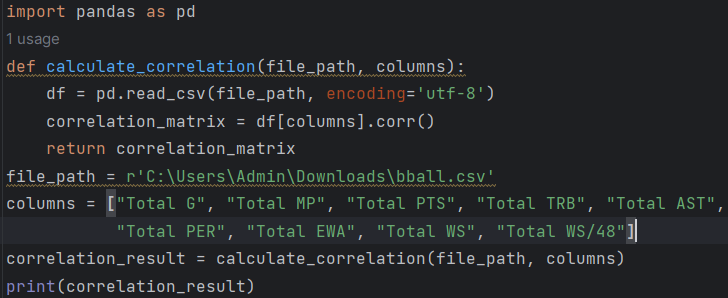
* là để các đồng thời:



* Số lượng dữ liệu là:



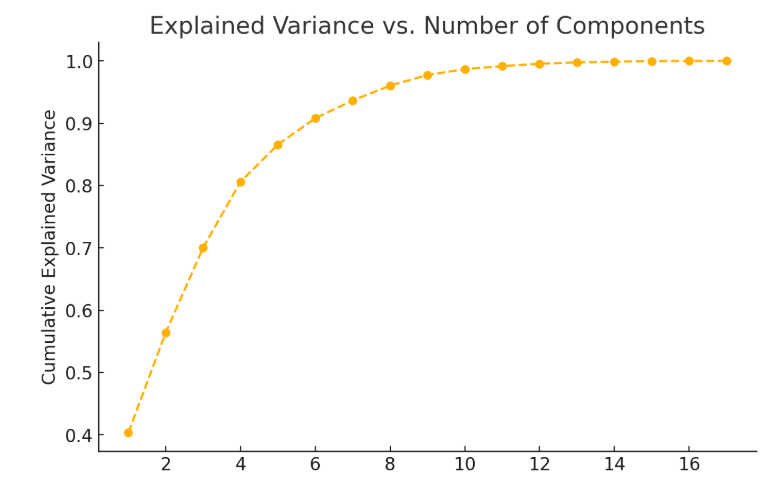
* & l#224;m cho m#7897;t c#225;i n#224;y.



* Không tìm thấy liên kết hầu hết liên kết dính cặp cặp của mỗi tập tin cài đặt dữ liệu. Ghi chú về các kết quả khi khởi chạy khi khởi chạy
* Tổng PTS và Tổng MP bị thiếu bởi vì những người chơi nhiều phút có xu hướng ghi điểm hơn.
* Tổng hồ sơ và Tổng AST cũng được hỗ trợ cho những người tham gia trong việc hỗ trợ và hỗ trợ.
* Có dữ liệu nào bị mất không?
* Như đã đề cập ở trên, không có bất kỳ dữ liệu còn thiếu nào trong bộ dữ liệu này.

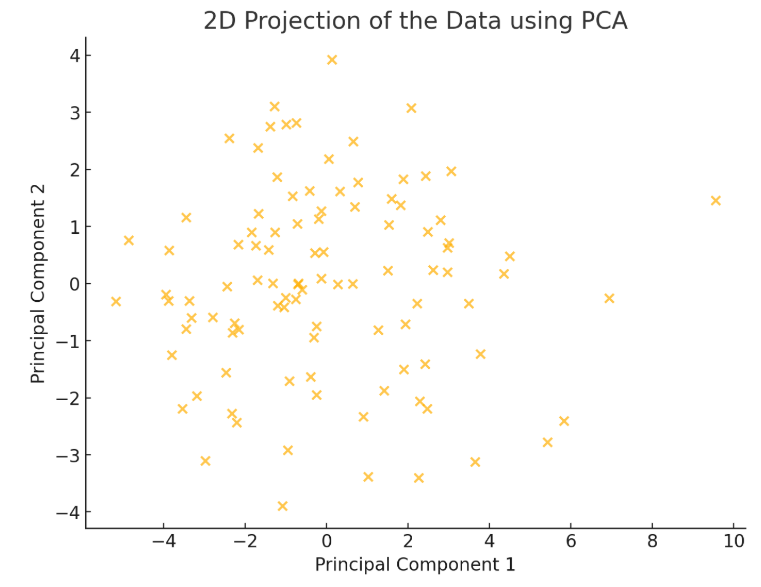
1. Không có

* Áp dụng PCA trên hai bộ dữ liệu đã chọn. Mô tả làm thế nào bạn chọn các thành phần chính. Có sự khác biệt nào trong khi sử dụng các thành phần có giá trị cao nhất và thấp nhất không?
* Hãy phân tích và phân tích con số các thành phần chính đã được sử dụng, phân tích và bình luận về kết quả đã đạt được.
* Dữ liệu sơ đồ phân tích trong 2D.

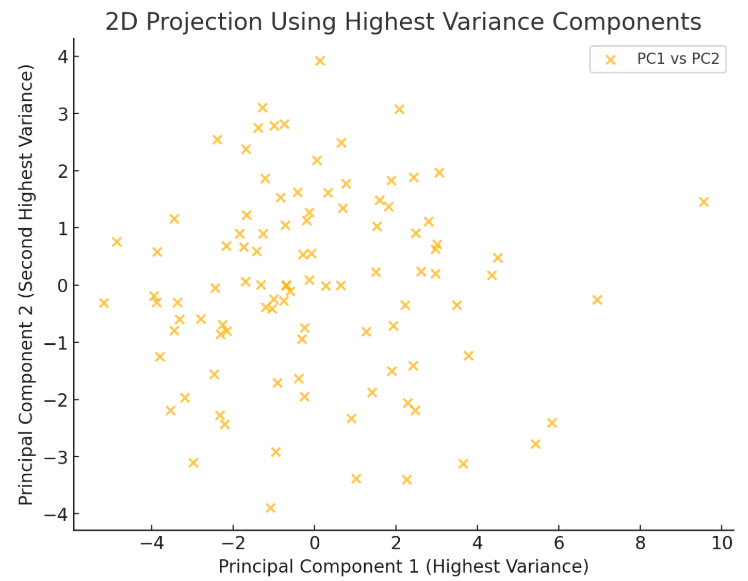
****

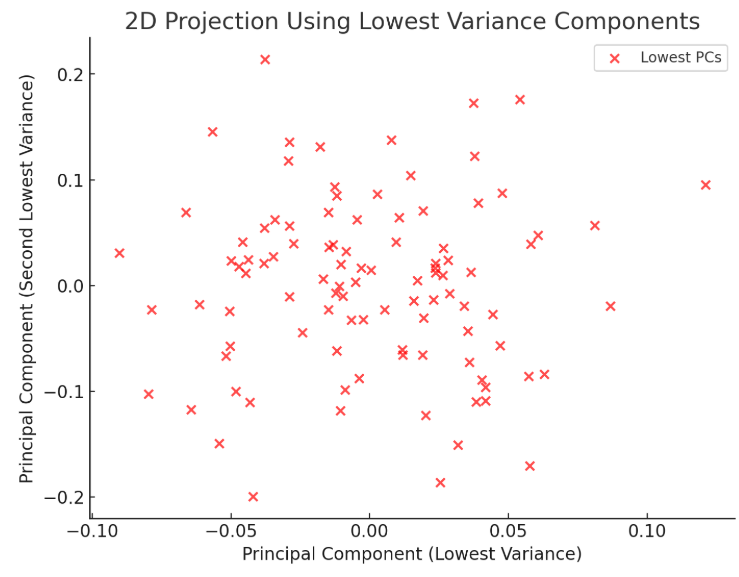
Nội dung hiển thị các tính năng xác suất giải thích là số lượng các số phần tử tăng.

Điều này giúp xác định bao nhiêu thành phần cần thiết để nắm bắt hầu hết các tính ưu tiên trong dữ liệu.

****

Các chi tiết phân loại đại diện cho sự phân phối dữ liệu trong 2 chiều bằng hai thành phần đầu tiên.





Nguyên nhân đầu tiên (dùng các thành phần chính yếu cao nhất) cho thấy một cấu trúc được sắp xếp chu đáo, chiếm hầu hết khả năng của bộ dữ liệu.

Âm mưu thứ hai (dùng các thành phần chính chứng hạn hẹp nhất) xuất hiện nhiều hơn và ồn ào hơn, vì những thành phần này chứa số lượng ít nhất có ý nghĩa. Giá trị này xác nhận rằng các thành phần chính tối đa là số nguyên tố chính tối đa cho việc phân tích và hoạt động.