**Ảnh có chứa quảng trường

Mô tả được tạo tự động** **ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

Ảnh có chứa văn bản, mẫu họa

Mô tả được tạo tự động

**BÁO CÁO ĐỒ ÁN**

**KHO DỮ LIỆU VÀ OLAP – IS217.O21**

**ĐỀ TÀI:**

**Phân tích dữ liệu về thông số trận đấu tại giải bóng rổ NBA ở Mỹ**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn**:  Ths.Đỗ Thị Minh Phụng | |
| **Sinh viên thực hiện:**  Nguyễn Trần Gia Kiệt – 21522258  Vương Thanh Linh – 21521082 | |
|  |  |

# LỜI CẢM ƠN

Trong cuộc sống của chúng ta, có lẽ ai cũng đã từng thất bại hoặc thành công, dù như thế nào thì đó cũng là kết quả nỗ lực của mỗi cá nhân cũng như tập thể. Và đằng sau đó chính là sự hỗ trợ giúp đỡ từ mọi người. Xét về mặt thành công, trong thực tế không có sự thành công nào mà không có sự giúp đỡ, nhất là trong học tập. Dân gian ta có câu “Không thầy đố mày làm nên” quả thật là đúng, học sinh không thể thành công nếu không có sự giúp đỡ, truyền đạt cũng như chỉ bảo tận tình của người Thầy. Hôm nay, để có thể hoàn thành được đồ án môn học này, nhóm chúng em rất biết ơn những thầy cô đã hỗ trợ tận tình, đã cung cấp cho chúng em nhiều kiến thức cũng như kinh nghiệm của người đi trước.

Lời đầu tiên, nhóm Nhóm 18 chúng em xin gửi lời cảm ơn chân thành đến tập thể quý Thầy Cô Trường Đại học Công Nghệ Thông Tin – Đại học Quốc Gia TP.HCM và quý Thầy Cô khoa Hệ thống thông tin đã giúp cho nhóm chúng em có những kiến thức cơ bản làm nền để thực hiện đề tài này.

Đặc biệt nhóm chúng em gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới Thạc sĩ Đỗ Thị Minh Phụng đã trực tiếp hướng dẫn tận tình, sửa chữa và đóng góp nhiều ý kiến quý báo giúp chúng em hoàn thành tốt báo cáo môn học của mình. Ngoài ra, chúng em cũng gửi lời cảm ơn đến tập thể lớp *IS217.O21* vì thời gian qua đã đồng hành cùng nhau. Cùng nhau học tập, tranh luận một cách sôi nổi để xây dựng bài học một cách tốt nhất. Cảm ơn các bạn đã cùng thảo luận, đánh giá và đóng góp ý kiến, cùng học hỏi nghiên cứu để thực hiện đồ án một cách tốt nhất có thể.

Cuối cùng, để đồ án được hoàn thành thì không thể nào cám ơn những người đã làm ra đó, cám ơn các bạn các thành viên trong nhóm đã chăm chỉ và chịu khó hoàn thành nhiệm vụ đúng tiến độ.

Trong thời gian một học kỳ thực hiện đề tài, nhóm chúng em đã vận dụng những kiến thức nền tảng đã tích lũy đồng thời kết hợp với việc học hỏi và nghiên cứu những kiến thức mới. Từ đó, nhóm chúng em vận dụng tối đa những gì đã thu thập được để hoàn thành một báo cáo đồ án tốt nhất. Tuy nhiên, trong quá trình thực hiện, nhóm chúng em không tránh khỏi những thiếu sót. Chính vì vậy, nhóm chúng em rất mong nhận được những sự đóng góp từ phía thầy cô nhằm hoàn thiện những kiến thức mà nhóm chúng em đã học tập và là hành trang để nhóm chúng em thực hiện tiếp các đề tài khác trong tương lai.

Một lần nữa, chúng em xin được gửi lời cảm ơn chân thành nhất của mình đến thầy, cô, chúng em chúc thầy, cô thật dồi dào sức khoẻ và thành công trên con đường giảng dạy của mình!

Thành phố Hồ Chí Minh, năm 2024

**Nhóm sinh viên thực hiện**

# NHẬN XÉT CỦA GIẢNG VIÊN

*……., ngày……...tháng……năm 2024*

**Người nhận xét**

*(Ký tên và ghi rõ họ tên****)***

# MỤC LỤC

[LỜI CẢM ƠN 1](#_Toc161600609)

[NHẬN XÉT CỦA GIẢNG VIÊN 3](#_Toc161600610)

[MỤC LỤC 4](#_Toc161600611)

[CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ KHO DỮ LIỆU 5](#_Toc161600612)

[1.1 Giới thiệu về dữ liệu 5](#_Toc161600613)

[1.1.1 Lý do chọn đề tài 5](#_Toc161600614)

[1.1.2 Giới thiệu dataset 5](#_Toc161600615)

[1.2 Mô tả các thuộc tính 6](#_Toc161600616)

[1.3 Thiết kế kho dữ liệu 8](#_Toc161600617)

[1.3.1 Sơ đồ bông tuyết minh hoạ 8](#_Toc161600618)

[1.3.2 Các bảng chiều 9](#_Toc161600619)

[1.3.3 Bảng sự kiện 9](#_Toc161600620)

[1.4 Câu truy vấn 9](#_Toc161600621)

# CHƯƠNG 1: GIỚI THIỆU TỔNG QUAN VỀ KHO DỮ LIỆU

## Giới thiệu về dữ liệu

### Lý do chọn đề tài

* **Sự phổ biến của NBA**: NBA là một trong những giải đấu bóng rổ nổi tiếng và phổ biến nhất trên toàn thế giới. Với sự hâm mộ rộng lớn, việc phân tích dữ liệu từ giải đấu này sẽ mang lại thông tin hữu ích và thú vị cho một đối tượng người đọc rất lớn.
* **Dữ liệu phong phú**: NBA cung cấp một lượng lớn dữ liệu từ mỗi trận đấu, bao gồm điểm số, thống kê cá nhân của cầu thủ, tỷ lệ thành công, và nhiều thông số khác. Điều này tạo điều kiện cho việc thực hiện phân tích sâu sắc để hiểu rõ hơn về các yếu tố ảnh hưởng đến kết quả của trận đấu.
* **Phân tích dữ liệu thể thao**: Bóng rổ là một trong những môn thể thao mà dữ liệu có thể được sử dụng rất hiệu quả để phân tích và dự đoán kết quả. Việc nghiên cứu về các mô hình dữ liệu trong bóng rổ có thể cung cấp thông tin giá trị cho các nhà quản lý đội bóng, người chơi, và cả những người cá cược.

### Giới thiệu dataset

* Tên dataset: NBA Traditional Boxscores 1997-2023
* Lần cập nhập gần nhất: 02 tháng 3 năm 2024
* Dataset bao gồm 29 cột thuộc tính và 65574 dòng dữ liệu chứa các điểm số của cầu thủ và đội từ mùa giải NBA 1996-97 đến 2022-23 được lấy trên web từ NBA.com.
* Dataset được một dataset expert có tên Szymon Jóźwiak cung cấp dựa vào dữ liệu trên NBA.com.
* Nguồn tải dataset: <https://www.kaggle.com/datasets/szymonjwiak/nba-traditional>

## Mô tả các thuộc tính

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Tên thuộc tính | Ý nghĩa | Kiểu dữ liệu |
|  | Gameid | NBA.com Game ID. | Long |
|  | Date | Ngày diễn ra trận đầu. | Date |
|  | Type | Loại của trận đấu. | Varchar |
|  | Teamid | NBA.com Team ID. | Long |
|  | Team | Tên đội bóng. | Varchar |
|  | Home | Đội bóng sân nhà. | Varchar |
|  | Away | Đội bóng sân khách. | Varchar |
|  | Min | Số phút thi đấu của một cầu thủ hoặc một đội. | Int |
|  | PTS | Số điểm đội bóng ghi được trong một trận đấu. | Int |
|  | FGM | Số lượng cú ném thành công mà cầu thủ hoặc đội làm được trong “field”. Bao gồm ném 2 điểm và ném 3 điểm. | Int |
|  | FGA | Số lần đội bóng hoặc cầu thủ cố gắng đưa bóng vào rổ từ bên trong các vùng “field” trong một trận đấu. | Int |
|  | FG% | Tỉ lệ ném thành công từ “field”. | Double |
|  | 3PM | Số bàn thắng 3 điểm mà cầu thủ hoặc đội ghi được trong một trận đấu. | Int |
|  | 3PA | Tổng số lần cầu thủ hoặc đội thực hiện cú ném 3 điểm, kể cả thành công hoặc không thành công. | Int |
|  | 3P% | Tỉ lệ thành công của các cú ném 3 điểm mà cầu thủ đội đó thực hiện | Double |
|  | FTM | Số quả ném phạt mà cầu thủ hoặc đội đó đã thực hiện và thành công. | Int |
|  | FTA | Số quả ném phạt mà cầu thủ hoặc đội đó đã thực hiện. | Int |
|  | FT% | Tỉ lệ thành công của các cú ném phạt được đội hoặc cầu thủ thực hiện. | Double |
|  | OREB | Số lần cầu thủ hoặc đội giành lại được bóng khi đội mình thực hiện cú ném không thành công từ vị trí tấn công. | Int |
|  | DREB | Số lần cầu thủ hoặc đội giành lại được bóng khi đội đối thủ thực hiện cú ném không thành công từ vị trí tấn công. | Int |
|  | REB | Tổng số lần Rebound cầu thủ hoặc đội bóng thực hiện được | Int |
|  | AST | Số lần kiến tạo của cầu thủ thuộc đội bóng đó thực hiện được | Int |
|  | TOV | Số lần đội mất quyền kiểm soát bóng vào tay đội đối thủ. | Int |
|  | STL | Số lần cầu thủ hoặc đội cướp được bóng từ đối thủ mà không gây ra lỗi. | Int |
|  | BLK | Số lần cầu thủ hoặc đội chặn cú ném từ đội đối thủ, ngăn bóng vào rổ | Int |
|  | PF | Số lỗi cá nhân mà cầu thủ hoặc đội đó vi phạm. | Int |
|  | +/- | “plus-minus” tổng số điểm đội mình ghi được trừ cho số điểm đội đối phương ghi được trong thời gian 2 đội thi đấu. | Int |
|  | Win | Đội chiến thắng (đội thắng sẽ là 1 còn đội thua là 0) | Int |
|  | season | Năm mùa giải diễn ra | Int |

## Thiết kế kho dữ liệu

### Sơ đồ bông tuyết minh hoạ

A close-up of a diagram

Description automatically generated

### Các bảng chiều

#### DimGame

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu | Ràng buộc | Ý nghĩa của thuộc tính |
|  | gameid | Long | Primary Key |  |
|  | matchDate | Date |  |  |
|  | 1stTeamId | long |  |  |
|  | 2ndTeamId | Long |  |  |
|  | 1stTeamPoint | Int |  |  |
|  | 2ndTeamPoint | Int |  |  |
|  | duration | Int |  |  |
|  | season | Int |  |  |

#### DimTime

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu | Ràng buộc | Ý nghĩa của thuộc tính |
|  | timekey | Int | Primary Key |  |
|  | gameId | Long | Foreign Key |  |
|  | day | Int |  |  |
|  | month | Int |  |  |
|  | year | Int |  |  |

#### DimGameType

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu | Ràng buộc | Ý nghĩa của thuộc tính |
|  | gameType | Varchar | Primary Key |  |
|  | typeName | Varchar |  |  |

#### DimTeam

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu | Ràng buộc | Ý nghĩa của thuộc tính |
|  | teamId | Long | Primary Key |  |
|  | teamName | Varchar |  |  |

### Bảng sự kiện

#### 1.3.3.1 FactGameStats

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| STT | Tên thuộc tính | Kiểu | Ràng buộc | Ý nghĩa của thuộc tính |
|  | statsId | Long | Primary Key |  |
|  | gameId | Long | Foreign Key |  |
|  | gameType | Varchar | Foreign Key |  |
|  | checkedTeamId |  |  |  |
|  | FGM | Int |  |  |
|  | FGA | Int |  |  |
|  | 3PM | Int |  |  |
|  | 3PA | Int |  |  |
|  | FTM | Int |  |  |
|  | FTA | Int |  |  |
|  | OREB | Int |  |  |
|  | BREB | Int |  |  |
|  | REB | Int |  |  |
|  | TOV | Int |  |  |
|  | STL | Int |  |  |
|  | BLK | Int |  |  |
|  | PF | Int |  |  |

## Câu truy vấn

1. Tổng số trận đấu của mỗi đội bóng?
2. Tổng số điểm ghi được bởi mỗi đội trong mùa giải?
3. Tổng số lần cản phá của mỗi đội?
4. Số lần cản phá cao nhất của mỗi đội trong một trận đấu?
5. Trung bình số lần cắt bóng của mỗi đội?
6. Tổng số lần cướp bóng của mỗi đội?
7. Trung bình số lần phát bóng thành công từ ngoài vòng 3 điểm của mỗi đội?
8. Trung bình số lần phát bóng thành công từ trong vòng 3 điểm của mỗi đội?