

A Minimal Book Example

Angelo Carmignan, Gabriel Bortol, Wesley Maia

2023-07-12

Contents

Preface	5
Sobre os Autores	7
1 Introdução	9
1.1 Objetivo	9
2 Dados	11
3 Exploração dos Dados (EDA)	13
3.1 Tratamento dos dados	13
4 Footnotes and citations	17
4.1 Footnotes	17
4.2 Citations	17
5 Blocks	19
5.1 Equations	19
5.2 Theorems and proofs	19
5.3 Callout blocks	19
6 Sharing your book	21
6.1 Publishing	21
6.2 404 pages	21
6.3 Metadata for sharing	21

Preface

In this portfolio, you will find case studies covering the area of Data Science.

Sobre os Autores

Chapter 1

Introdução

A NBA (National Basketball Association) é uma das ligas de basquete mais populares e prestigiadas do mundo, com uma rica história que se estende por 77 anos. Desde sua fundação em 1946, a NBA tem sido palco de inúmeras façanhas atléticas, rivalidades intensas e momentos memoráveis que cativaram os fãs de basquete em todo o mundo.

Neste trabalho de Visualização de Dados, exploraremos um conjunto abrangente de estatísticas dos últimos 71 anos da NBA. Utilizando o Jupyter Notebook, mergulharemos nesses dados para extrair insights valiosos sobre as equipes, jogadores e padrões que moldaram a liga ao longo das décadas.

1.1 Objetivo

O objetivo desta análise é investigar diversas facetas do basquete profissional, desde o desempenho das equipes até as estatísticas individuais dos jogadores. Por meio de técnicas de análise de dados e visualização, buscaremos responder a perguntas como:

Quais equipes dominaram a NBA ao longo dos anos? Quais jogadores tiveram as melhores performances estatísticas em diferentes épocas? Existem tendências ou padrões significativos nas estatísticas da NBA ao longo das décadas? Como o jogo evoluiu em termos de estilo de jogo, pontuação média e estilos de arremesso? Ao responder a essas perguntas, esperamos obter uma compreensão mais profunda da evolução da NBA e das dinâmicas que impulsionam o sucesso das equipes e dos jogadores ao longo do tempo. Esses insights não apenas fornecerão informações interessantes sobre a história da liga, mas também poderão ajudar a prever tendências futuras e orientar estratégias para equipes e jogadores no presente.

Chapter 2

Dados

O projeto tem um conjunto de dados fornecido pelo Kaggle chamado nba.csv. NA base apresenta os dados dos jogadores de todas as temporadas de 1951 a 2022, com um total de 33330 ocorrências

2.0.1 Descrição da base

As colunas são descritas a seguir:

POR JOGADOR:

Variável	Descrição
Rank	A classificação do jogador (ordenado por pontos marcados a cada temporada)
Year	O ano da temporada (por exemplo, “2018-19”)
Season	O ano de início da temporada (por exemplo, 2018)
Start	
Year	
Season	Temporada regular ou playoffs
Type	
Player	Um ID gerado para cada jogador
ID	
Player	O nome do jogador
Team	ID gerado para cada equipe
ID	
Team	A equipe do jogador na respectiva temporada
Games	Jogos disputados na respectiva temporada
Played	
Minutes	Minutos jogados na respectiva temporada
Played	

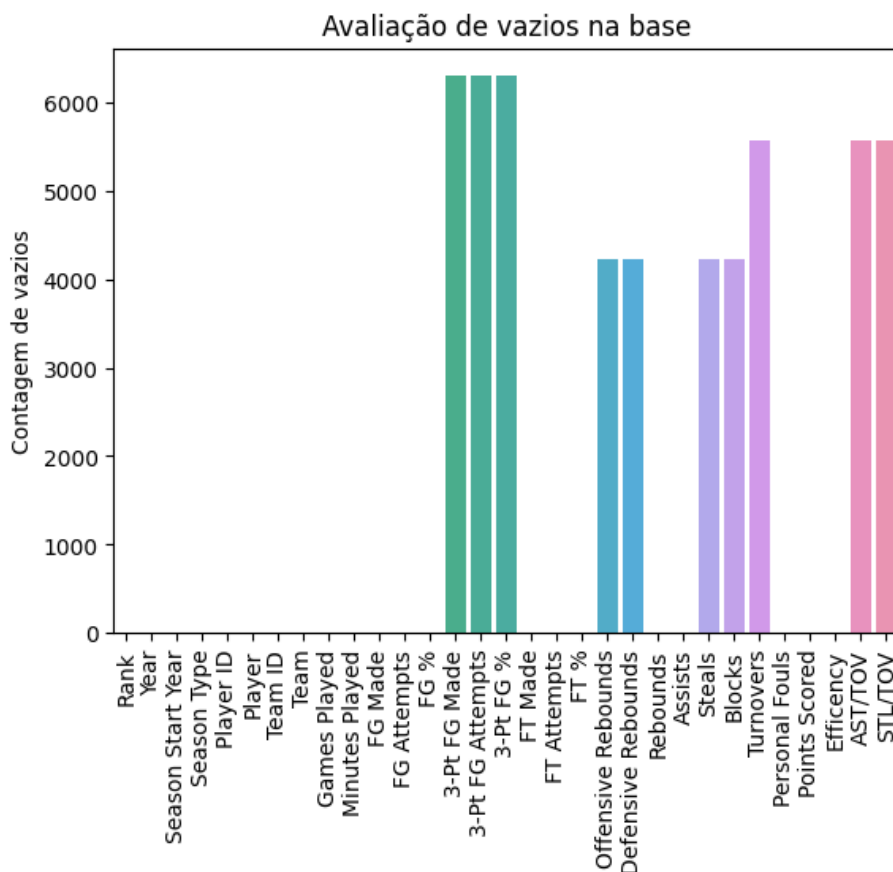
Variável	Descrição
FG Made	Cestas de campo convertidas (Field Goals Made)
FG At-tempts	Tentativas de cestas de campo (Field Goals Attempted)
FG %	Porcentagem de acertos de cestas de campo (Field Goal Percentage)
3-Pt FG Made	Cestas de três pontos convertidas (3 Point Field Goals Made)
3-Pt FG At-tempts	Tentativas de cestas de três pontos (3 Point Field Goals Attempted)
3-Pt FG %	Porcentagem de acertos de cestas de três pontos (3 Point Field Goal Percentage)
FT Made	Lances livres convertidos (Free Throws Made)
FT At-tempts	Tentativas de lances livres (Free Throws Attempted)
FT %	Porcentagem de acertos de lances livres (Free Throw Percentage)
Offensive Rebounds	Rebotes ofensivos
Defensive Rebounds	Rebotes defensivos
Total Rebounds	Total de rebotes (ofensivos + defensivos)
Assists	Assistências
Steals	Roubos de bola
Blocks	Tocos (bloqueios de arremessos)
Turnovers	Perdas de bola (erros)
Personal Fouls	Faltas pessoais
Points Scored	Pontos marcados
Efficiency	Eficiência calculada como $(\text{Pontos Marcados} + \text{Rebotes} + \text{Assistências} + \text{Roubos de Bola} + \text{Tocos} - \text{Chutes de Campo Perdidos} - \text{Lances Livres Perdidos} - \text{Perdas de Bola})$ dividido por Jogos Disputados
AST/TOV	Taxa de assistências para turnovers (Assist-to-Turnover ratio)
STL/TOV	Taxa de roubos de bola para turnovers (Steal-to-Turnover ratio)

Chapter 3

Exploração dos Dados (EDA)

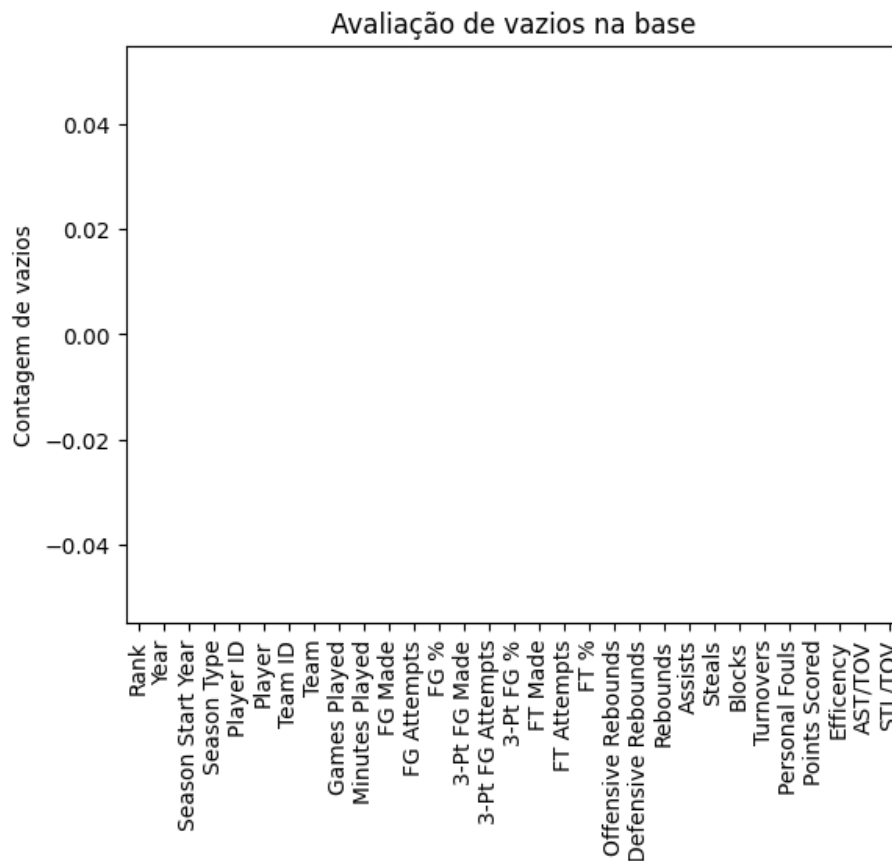
3.1 Tratamento dos dados

Vimos que existem alguns valores vazios o que é um problema para as análises que serão feitas. Vamos conferir quantas linhas do dataset existem desse tipo de dados.



O DataFrame apresenta valores nulos em algumas colunas, como 3-Pt FG Made, 3-Pt FG Attempts, 3-Pt FG %, Offensive Rebounds, Defensive Rebounds, Steals, Blocks, Turnovers, AST/TOV e STL/TOV. Esses valores nulos podem indicar a ausência de dados ou informações faltantes para algumas estatísticas específicas dos jogadores em determinadas temporadas.

Para manter a consistência e garantir a confiabilidade da análise, optou-se por filtrar o DataFrame, excluindo as temporadas anteriores a 1978. Dessa forma, as colunas mencionadas estarão preenchidas a partir desse ano, permitindo uma análise mais completa e precisa das estatísticas dos jogadores da NBA. Essa decisão foi tomada para evitar distorções nos resultados devido à ausência de dados em períodos anteriores, garantindo que a análise seja baseada em informações mais completas e recentes.



3.1.1 Primeira pergunta: O que define um jogador bom ?

Para simplificação vamos utilizar o rank apresentado no dataset, que seriam os jogadores ordenados de acordo com a pontuação por temporada.

Apesar dessa rank não levar em consideração fatores defensivos, será feito uma avaliação se os maiores “cestinhas” também apresentam características defensivas acima da média.

A primeiro momento vamos avaliar como as informações estatísticas de cada jogador por temporada varia e como está relacionado com o rank. Para tanto será utilizado a técnica do PCA para entender melhor esse comportamento.

Chapter 4

Footnotes and citations

4.1 Footnotes

Footnotes are put inside the square brackets after a caret `^[]`. Like this one ¹.

4.2 Citations

Reference items in your bibliography file(s) using `@key`.

For example, we are using the **bookdown** package [Xie, 2023] (check out the last code chunk in `index.Rmd` to see how this citation key was added) in this sample book, which was built on top of R Markdown and **knitr** [Xie, 2015] (this citation was added manually in an external file `book.bib`). Note that the `.bib` files need to be listed in the `index.Rmd` with the YAML `bibliography` key.

The RStudio Visual Markdown Editor can also make it easier to insert citations: <https://rstudio.github.io/visual-markdown-editing/#/citations>

¹This is a footnote.

Chapter 5

Blocks

5.1 Equations

Here is an equation.

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (5.1)$$

You may refer to using `\@ref{eq:binom}`, like see Equation (5.1).

5.2 Theorems and proofs

Labeled theorems can be referenced in text using `\@ref{thm:tri}`, for example, check out this smart theorem 5.1.

Theorem 5.1. *For a right triangle, if c denotes the length of the hypotenuse and a and b denote the lengths of the **other** two sides, we have*

$$a^2 + b^2 = c^2$$

Read more here <https://bookdown.org/yihui/bookdown/markdown-extensions-by-bookdown.html>.

5.3 Callout blocks

The R Markdown Cookbook provides more help on how to use custom blocks to design your own callouts: <https://bookdown.org/yihui/rmarkdown-cookbook/custom-blocks.html>

Chapter 6

Sharing your book

6.1 Publishing

HTML books can be published online, see: <https://bookdown.org/yihui/bookdown/publishing.html>

6.2 404 pages

By default, users will be directed to a 404 page if they try to access a webpage that cannot be found. If you'd like to customize your 404 page instead of using the default, you may add either a `_404.Rmd` or `_404.md` file to your project root and use code and/or Markdown syntax.

6.3 Metadata for sharing

Bookdown HTML books will provide HTML metadata for social sharing on platforms like Twitter, Facebook, and LinkedIn, using information you provide in the `index.Rmd` YAML. To setup, set the `url` for your book and the path to your `cover-image` file. Your book's `title` and `description` are also used.

This `gitbook` uses the same social sharing data across all chapters in your book—all links shared will look the same.

Specify your book's source repository on GitHub using the `edit` key under the configuration options in the `_output.yml` file, which allows users to suggest an edit by linking to a chapter's source file.

Read more about the features of this output format here:

<https://pkgs.rstudio.com/bookdown/reference/gitbook.html>

Or use:

```
?bookdown::gitbook
```

Bibliography

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL <http://yihui.org/knitr/>. ISBN 978-1498716963.

Yihui Xie. *bookdown: Authoring Books and Technical Documents with R Markdown*, 2023. URL <https://CRAN.R-project.org/package=bookdown>. R package version 0.34.