Análise descritiva e estatística da base dos jogadores ativos da NBA

Marcos Perazo Viana

12 abril, 2025

Introdução

Este projeto utiliza como base de dados informações sobre jogadores ativos da NBA, obtidas no site oficial da NBA, na aba *Players*, no dia 22 de março de 2025. Os dados coletados incluem: nomes dos jogadores, equipes em que atuam, posições, altura, peso, data de nascimento, idade, média de pontos, rebotes e assistências por jogo da atual temporada, além de informações sobre o *Draft* e experiência profissional. Adicionalmente, foi criada uma variável indicando o mês de nascimento dos jogadores. Com esses dados, é possível investigar diversas correlações estatísticas relevantes que ajudam a compreender o desempenho e as características dos atletas. Por exemplo, pode-se analisar a relação entre altura dos jogadores e a média de rebotes por jogo. Partindo da hipótese de que jogadores mais altos possuem vantagem nesse quesito, essa análise busca confirmar ou refutar tal tendência. Outras correlações potencialmente interessantes incluem:

- Peso e média de pontos por jogo: Investiga se jogadores com maior massa corporal têm vantagens em marcar pontos, especialmente em posições como pivô.
- Assistências por jogo e pontos por jogo: Examina a relação entre jogadores que criam oportunidades para seus colegas de equipe e aqueles que executam as finalizações.
- Altura e assistências por jogo: Avalia se jogadores mais baixos, como armadores, têm maior tendência a contribuir com assistências. Média de pontos por jogo e rebotes por jogo: Analisa a relação entre a capacidade de pontuar e pegar rebotes, considerando as diferentes posições em quadra.

Essas análises têm como objetivo explorar padrões e características que podem enriquecer a compreensão sobre o desempenho dos jogadores da NBA.

Carregamento da base.

```
## Rows: 569 Columns: 14
## -- Column specification ------
## Delimiter: ","
## chr (10): Equipe, Posicao, Nome, PPG, RPG, APG, País, Idade, DRAFT, Experiê...
## dbl (3): Altura, Peso, Mês
## date (1): Data_de_nascimento
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
## Warning: There were 3 warnings in 'mutate()'.
## The first warning was:
## i In argument: 'PPG = as.numeric(PPG)'.
## Caused by warning:
## ! NAs introduzidos por coerção
## i Run 'dplyr::last_dplyr_warnings()' to see the 2 remaining warnings.
```

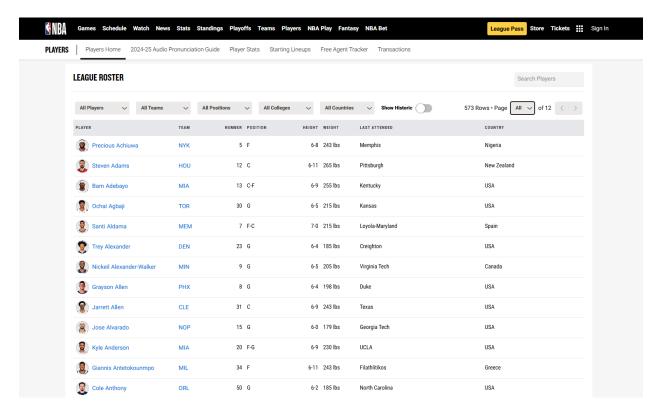


Figure 1: Site da NBA

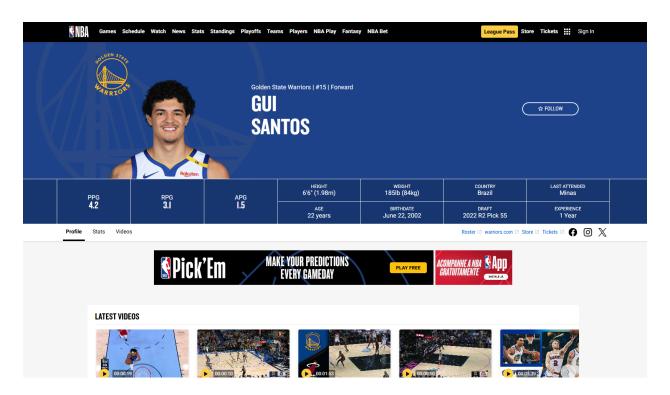


Figure 2: Gui Santos

Observações iniciais:

carefully

- 1 Fazendo uma análise preliminar, verificamos que a linha 507 não possui dados em nenhuma das colunas, e foi determinada a eliminação dessa linha.
- 2 Após verificar as estatisticas individuais dos atletas que receberam NA nas colunas PPG, APG e RPG, verificamos que esses campos deveriam receber valor zero. O que foi realizado.

568.0000000

568.0000000

100.0000000

568.0000000

568.0000000

100.0000000

```
## Warning in parse_call(mc = match.call(), var_name = (ncol(xx) == 1), var_label
## = (ncol(xx) == : metadata extraction terminated unexpectedly; inspect results
```

Error in table(names(candidates))[["tested"]]: indice fora dos limites

	Altura	RPG
Mean	1.9949296	3.4693662
Std.Dev	0.0794136	2.4857786
Min	1.7300000	0.0000000
Q1	1.9300000	1.6000000
Median	1.9800000	3.1000000
Q3	2.0600000	4.6000000
Max	2.2400000	13.9000000
MAD	0.0741300	2.2239000
IQR	0.1300000	3.0000000
CV	0.0398077	0.7164936
Skewness	0.0603177	1.2213483
SE.Skewness	0.1025083	0.1025083
Kurtosis	-0.2866479	1.7838539

N.Valid

Pct.Valid

Ν