

Projeto_NBA

Thomas Raphael Zonta

2025-04-01

Importando bibliotecas

```
library(readxl)
library(tidyverse)
library(ggplot2)
library(dlookr)
library(dplyr)
library(purrr)
```

Importando base dos jogos

```
jogos_2024 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogos_2024.xlsx")
```

Importando base jogadores

```
jogadores_2020 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogadores_2020.xlsx")
jogadores_2021 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogadores_2021.xlsx")
jogadores_2022 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogadores_2022.xlsx")
jogadores_2023 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogadores_2023.xlsx")
jogadores_2024 <- read_excel("C:/Users/Pichau/Documents/RSTUDIO/NBA/jogadores_2024.xlsx")
```

Checando cabeçalho base jogos

Para poder visualizar quais dados nós temos para trabalhar usamos o comando `head()` para visualizar o topo do dataset.

```
head(jogos_2024)
```

```
## # A tibble: 6 x 6
##   'Round Number' Date           Location   'Home Team' 'Away Team' Result
##             <dbl> <dtm>         <chr>      <chr>      <chr>      <chr>
```

```
## 1          1 2024-10-22 23:30:00 TD Garden    Boston Cel~ New York K~ 132 --
## 2          1 2024-10-23 02:00:00 Crypto.com ~ Los Angele~ Minnesota ~ 110 --
## 3          1 2024-10-23 23:00:00 Little Caes~ Detroit Pi~ Indiana Pa~ 109 --
## 4          1 2024-10-23 23:30:00 State Farm ~ Atlanta Ha~ Brooklyn N~ 120 --
## 5          1 2024-10-23 23:30:00 Kaseya Cent~ Miami Heat  Orlando Ma~ 97  - ~
## 6          1 2024-10-23 23:30:00 Wells Fargo~ Philadelph~ Milwaukee ~ 109 --
```

Separar resultados dos placares existentes

Iremos separar os resultados dos jogos para que possamos analisa-los individualmente no cálculo de probabilidade.

```
jogos_2024 <- jogos_2024 %>%
  separate(Result, into = c("Placar_Casa", "Placar_Fora"), sep= " - ", convert = TRUE)
```

Renomear dataset

Renomearemos as colunas para facilitar o entendimento e mudaremos o tipo de dado para int nas colunas que foram criadas.

```
colnames(jogos_2024) <- c("Numero_Round", "Data", "Local", "Time_Casa", "Visitante", "Placar_Casa", "Placar_Fora")

jogos_2024 <- jogos_2024 %>%
  mutate(
    Placar_Casa = as.numeric(Placar_Casa),
    Placar_Fora = as.numeric(Placar_Fora)
  )

head(jogos_2024)
```

```
## # A tibble: 6 x 7
##   Numero_Round Data          Local          Time_Casa Visitante Placar_Casa
##           <dbl> <dtm>          <chr>          <chr>      <chr>          <dbl>
## 1             1 2024-10-22 23:30:00 TD Garden    Boston C~ New York~          132
## 2             1 2024-10-23 02:00:00 Crypto.com A~ Los Ange~ Minnesot~          110
## 3             1 2024-10-23 23:00:00 Little Caesa~ Detroit ~ Indiana ~          109
## 4             1 2024-10-23 23:30:00 State Farm A~ Atlanta ~ Brooklyn~          120
## 5             1 2024-10-23 23:30:00 Kaseya Center Miami He~ Orlando ~           97
## 6             1 2024-10-23 23:30:00 Wells Fargo ~ Philadel~ Milwauke~          109
## # i 1 more variable: Placar_Fora <dbl>
```

Estatísticas dos times

Iremos calcular as estatísticas dos times separando por jogos em casa e fora.

```
estatisticas_times <- jogos_2024 %>%
  filter(!is.na(Placar_Casa) & !is.na(Placar_Fora)) %>%
  group_by(Time_Casa) %>%
```

```

summarise(
  Media_Pontos_Casa = mean(Placar_Casa, na.rm = TRUE),
  Media_Pontos_Sofridos_Casa = mean(Placar_Fora, na.rm = TRUE)
) %>%
rename(Time = Time_Casa) %>%
full_join(
  jogos_2024 %>%
    filter(!is.na(Placar_Casa) & !is.na(Placar_Fora)) %>%
    group_by(Visitante) %>%
    summarise(
      Media_Pontos_Fora = mean(Placar_Fora, na.rm = TRUE),
      Media_Pontos_Sofridos_Fora = mean(Placar_Casa, na.rm = TRUE)
    ) %>%
    rename(Time = Visitante),
  by = "Time"
)

head(estatisticas_times)

```

```

## # A tibble: 6 x 5
##   Time                Media_Pontos_Casa Media_Pontos_Sofrido~1 Media_Pontos_Fora
##   <chr>                <dbl>                <dbl>                <dbl>
## 1 Atlanta Hawks        117.                119.                117.
## 2 Boston Celtics       117.                109.                116.
## 3 Brooklyn Nets        104.                111.                107.
## 4 Charlotte Hornets     106.                111.                105.
## 5 Chicago Bulls         116.                120.                118.
## 6 Cleveland Cavaliers   123.                110.                122.
## # i abbreviated name: 1: Media_Pontos_Sofridos_Casa
## # i 1 more variable: Media_Pontos_Sofridos_Fora <dbl>

```

Associar estatísticas dos próximos jogos

Vamos associar as estatísticas do próximos jogos do time da casa e visitantes.

```

jogos_futuros_tmp <- jogos_2024 %>%
  filter(is.na(Placar_Casa) & is.na(Placar_Fora)) %>%
  left_join(estatisticas_times, by = c("Time_Casa" = "Time"))

colnames(jogos_futuros_tmp)

```

```

## [1] "Numero_Round"      "Data"
## [3] "Local"             "Time_Casa"
## [5] "Visitante"         "Placar_Casa"
## [7] "Placar_Fora"       "Media_Pontos_Casa"
## [9] "Media_Pontos_Sofridos_Casa" "Media_Pontos_Fora"
## [11] "Media_Pontos_Sofridos_Fora"

```

```

jogos_futuros_tmp2 <- jogos_futuros_tmp %>%
  left_join(estatisticas_times, by = c("Visitante" = "Time"))

```

```
head(jogos_futuros_tmp2)
```

```
## # A tibble: 6 x 15
##   Numero_Round Data          Local          Time_Casa Visitante Placar_Casa
##   <dbl> <dtm>          <chr>          <chr>      <chr>      <dbl>
## 1      20 2025-03-20 23:00:00 Spectrum Cen~ Charlott~ New York~      NA
## 2      20 2025-03-20 23:00:00 Gainbridge F~ Indiana ~ Brooklyn~      NA
## 3      20 2025-03-21 02:00:00 Chase Center Golden S~ Toronto ~      NA
## 4      20 2025-03-21 02:00:00 Golden 1 Cen~ Sacramen~ Chicago ~      NA
## 5      20 2025-03-21 02:30:00 Crypto.com A~ Los Ange~ Milwauke~      NA
## 6      20 2025-03-21 23:00:00 Capital One ~ Washingt~ Orlando ~      NA
## # i 9 more variables: Placar_Fora <dbl>, Media_Pontos_Casa.x <dbl>,
## #   Media_Pontos_Sofridos_Casa.x <dbl>, Media_Pontos_Fora.x <dbl>,
## #   Media_Pontos_Sofridos_Fora.x <dbl>, Media_Pontos_Casa.y <dbl>,
## #   Media_Pontos_Sofridos_Casa.y <dbl>, Media_Pontos_Fora.y <dbl>,
## #   Media_Pontos_Sofridos_Fora.y <dbl>
```

Renomear colunas

Vamos renomear as colunas e eliminar as colunas com valores duplicados.

```
jogos_futuros <- jogos_futuros_tmp2 %>%
  rename(
    Media_Pontos_Casa_C = Media_Pontos_Casa.x,
    Media_Pontos_Sofridos_C = Media_Pontos_Sofridos_Casa.x,
    Media_Pontos_Fora_V = Media_Pontos_Fora.y,
    Media_Pontos_Sofridos_V = Media_Pontos_Sofridos_Fora.y
  ) %>%
  select(-Media_Pontos_Casa.y, -Media_Pontos_Sofridos_Casa.y, -Media_Pontos_Fora.x, -Media_Pontos_Sofridos_Fora.x)

colnames(jogos_futuros)
```

```
## [1] "Numero_Round"      "Data"
## [3] "Local"             "Time_Casa"
## [5] "Visitante"         "Placar_Casa"
## [7] "Placar_Fora"       "Media_Pontos_Casa_C"
## [9] "Media_Pontos_Sofridos_C" "Media_Pontos_Fora_V"
## [11] "Media_Pontos_Sofridos_V"
```

Calcular ataque e defesa

Iremos calcular com base nos resultados dos jogos até o momento o poder ofensivo e defensivo dos times

```
estatisticas_times <- estatisticas_times %>%
  mutate(
    Forca_Ofensiva = (Media_Pontos_Casa + Media_Pontos_Fora) / 2,
    Forca_Defensiva = (Media_Pontos_Sofridos_Casa + Media_Pontos_Sofridos_Fora) / 2
  )
```

```
head(estatisticas_times)
```

```
## # A tibble: 6 x 7
##   Time                Media_Pontos_Casa Media_Pontos_Sofrido~1 Media_Pontos_Fora
##   <chr>                <dbl>          <dbl>          <dbl>
## 1 Atlanta Hawks        117.          119.          117.
## 2 Boston Celtics        117.          109.          116.
## 3 Brooklyn Nets        104.          111.          107.
## 4 Charlotte Hornets     106.          111.          105.
## 5 Chicago Bulls         116.          120.          118.
## 6 Cleveland Cavaliers   123.          110.          122.
## # i abbreviated name: 1: Media_Pontos_Sofridos_Casa
## # i 3 more variables: Media_Pontos_Sofridos_Fora <dbl>, Forca_Ofensiva <dbl>,
## #   Forca_Defensiva <dbl>
```

Calculo de pontos esperados

Iremos calcular os pontos esperados baseado na quantidade de pontos que o time marca em casa + pontos que o adversário sofre fora /2 e semelhante para os visitantes.

```
jogos_futuros <- jogos_futuros %>%
  mutate(
    Pontos_Esperados_Casa = (Media_Pontos_Casa_C + Media_Pontos_Sofridos_V) / 2,
    Pontos_Esperados_Fora = (Media_Pontos_Fora_V + Media_Pontos_Sofridos_C) / 2
  )

head(jogos_futuros)
```

```
## # A tibble: 6 x 13
##   Numero_Round Data          Local          Time_Casa Visitante Placar_Casa
##   <dbl> <dtm>          <chr>          <chr>      <chr>      <dbl>
## 1      20 2025-03-20 23:00:00 Spectrum Cen~ Charlott~ New York~      NA
## 2      20 2025-03-20 23:00:00 Gainbridge F~ Indiana ~ Brooklyn~      NA
## 3      20 2025-03-21 02:00:00 Chase Center Golden S~ Toronto ~      NA
## 4      20 2025-03-21 02:00:00 Golden 1 Cen~ Sacramen~ Chicago ~      NA
## 5      20 2025-03-21 02:30:00 Crypto.com A~ Los Ange~ Milwauke~      NA
## 6      20 2025-03-21 23:00:00 Capital One ~ Washingt~ Orlando ~      NA
## # i 7 more variables: Placar_Fora <dbl>, Media_Pontos_Casa_C <dbl>,
## #   Media_Pontos_Sofridos_C <dbl>, Media_Pontos_Fora_V <dbl>,
## #   Media_Pontos_Sofridos_V <dbl>, Pontos_Esperados_Casa <dbl>,
## #   Pontos_Esperados_Fora <dbl>
```

Probabilidade de vitória

Iremos calcular a probabilidade de vitória do time da casa e do visitante

```
jogos_futuros <- jogos_futuros %>%
  mutate(
    Prob_Vitoria_Casa = Pontos_Esperados_Casa / (Pontos_Esperados_Casa + Pontos_Esperados_Fora),
```

```

    Prob_Vitoria_Visitante = 1 - Prob_Vitoria_Casa
  )

head(jogos_futuros)

## # A tibble: 6 x 15
##   Numero_Round Data          Local          Time_Casa Visitante Placar_Casa
##   <dbl> <dtm>          <chr>          <chr>      <chr>      <dbl>
## 1      20 2025-03-20 23:00:00 Spectrum Cen~ Charlott~ New York~      NA
## 2      20 2025-03-20 23:00:00 Gainbridge F~ Indiana ~ Brooklyn~      NA
## 3      20 2025-03-21 02:00:00 Chase Center  Golden S~ Toronto ~      NA
## 4      20 2025-03-21 02:00:00 Golden 1 Cen~ Sacramen~ Chicago ~      NA
## 5      20 2025-03-21 02:30:00 Crypto.com A~ Los Ange~ Milwauke~      NA
## 6      20 2025-03-21 23:00:00 Capital One ~ Washingt~ Orlando ~      NA
## # i 9 more variables: Placar_Fora <dbl>, Media_Pontos_Casa_C <dbl>,
## #   Media_Pontos_Sofridos_C <dbl>, Media_Pontos_Fora_V <dbl>,
## #   Media_Pontos_Sofridos_V <dbl>, Pontos_Esperados_Casa <dbl>,
## #   Pontos_Esperados_Fora <dbl>, Prob_Vitoria_Casa <dbl>,
## #   Prob_Vitoria_Visitante <dbl>

```

Tabela de Resultados previstos

Agora criamos a tabela de resultados previstos seguindo a ordem original da tabela baseado nos calculos que realizamos e também calculamos e adicionamos os placares para cada jogo.

```

jogos_futuros <- jogos_futuros %>%
  mutate(
    Placar_Estimado_Casa = round(Media_Pontos_Casa_C),
    Placar_Estimado_Fora = round(Media_Pontos_Fora_V)
  )

tabela_resultados_previstos <- jogos_futuros %>%
  select(Time_Casa, Visitante, Prob_Vitoria_Casa, Prob_Vitoria_Visitante, Placar_Estimado_Casa, Placar_Estimado_Fora)

# Exibir a tabela
tabela_resultados_previstos

```

```

## # A tibble: 193 x 6
##   Time_Casa          Visitante Prob_Vitoria_Casa Prob_Vitoria_Visitante
##   <chr>          <chr>          <dbl>          <dbl>
## 1 Charlotte Hornets New York Kni~      0.492          0.508
## 2 Indiana Pacers   Brooklyn Nets      0.510          0.490
## 3 Golden State Warriors Toronto Rapt~      0.511          0.489
## 4 Sacramento Kings  Chicago Bulls      0.505          0.495
## 5 Los Angeles Lakers Milwaukee Bu~      0.509          0.491
## 6 Washington Wizards Orlando Magic      0.491          0.509
## 7 Miami Heat       Houston Rock~      0.497          0.503

```

```
## 8 Minnesota Timberwolves New Orleans ~ 0.519 0.481
## 9 Oklahoma City Thunder Charlotte Ho~ 0.528 0.472
## 10 San Antonio Spurs Philadelphia~ 0.506 0.494
## # i 183 more rows
## # i 2 more variables: Placar_Estimado_Casa <dbl>, Placar_Estimado_Fora <dbl>
```

```
#write.csv(tabela_resultados_previstos, "resultado_previstos_att.csv", row.names = FALSE)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.