

# **MATRIZ DE CORRELAÇÃO ENTRE AS QUATRO PRINCIPAIS LIGAS DE ESPORTES DOS ESTADOS UNIDOS**

GUILHERME DE MORAES MASUKO

2022

## **RESUMO**

**Palavras-chave:** Correlação. Esporte.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>3</b>
<b>2</b>	<b>METODOLOGIA . . . . .</b>	<b>4</b>
<b>3</b>	<b>DADOS . . . . .</b>	<b>6</b>
<b>4</b>	<b>CONCLUSÃO . . . . .</b>	<b>9</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>10</b>

## **1 INTRODUÇÃO**

Os Estados Unidos é um país conhecido pelo seu ótimo desempenho no esporte. Entre os esportes mais populares, estão o futebol americano, basquete, hóquei e beisebol. O Super Bowl é

## 2 METODOLOGIA

Para mensurar a relação entre a performance média das regiões metropolitanas nas ligas norte-americanas NBA, MLB e NFL, será proposto como medidor de performance, visualizado na equação 2.1, a razão entre vitórias e jogos de cada time, ou seja

$$P_i = \frac{W_i}{G_i} \quad (2.1)$$

onde  $P_i$  é a performance do time  $i$ ,  $W_i$  é a quantidade de vitórias na temporada regular do time  $i$  e  $G_i$  é a quantidade de jogos da temporada regular do time  $i$ .

Para a liga NHL, como há um esquema de pontuação diferente, onde o time ganha dois pontos caso vença, um ponto caso seja derrotado em um overtime, e zero caso seja derrotado no tempo normal, o medidor de performance é proposto a partir da razão entre pontuação do time e pontuação possível, este último travado em 164 (pontuação caso um time ganhar as 82 partidas que contém uma temporada regular). A equação 2.2 ilustra isso.

$$P_i = \frac{Pt_i}{\max(Pt)} = \frac{Pt_i}{164} \quad (2.2)$$

onde  $P_i$  é a performance do time  $i$ ,  $Pt_i$  é a pontuação na temporada regular do time  $i$  e  $\max(Pt)$  é a pontuação máxima, de 164 pontos, da temporada regular que um time pode atingir.

Possivelmente vários times podem estar situados em mesma região metropolitana e para chegar na performance da região será utilizado a média simples da performance de cada time.

E, finalmente, para mensurar a relação entre a performance média das regiões metropolitanas usaremos a matriz de correlação para as quatro principais ligas norte-americanas. Como pode ser visto em Wikipédia (2021a), correlação é uma medida estatística de relação (causal ou não-causal) entre duas variáveis abordando o comportamento da relação entre elas. O coeficiente de correlação utilizado pela presente análise será o de Pearson, o coeficiente de correlação mais comumente usado.

O coeficiente de correlação de Pearson, segundo Wikipédia (2021b), define o quanto e em qual direção duas variáveis estão relacionadas, sendo obtida a partir da seguinte fórmula demonstrada na equação 2.3 abaixo.. Usado normalmente como  $\rho$ , o coeficiente pode ter valores em um range de  $-1$  à  $1$ , isto é,  $-1 \leq \rho \leq 1$ , onde  $\rho = -1$  significa uma relação perfeitamente negativa entre as variáveis,  $\rho = 1$  uma relação perfeitamente positiva, e  $\rho = 0$  uma relação de não dependência linear (isso não diz nada sobre não haver dependência de modo geral entre as variáveis).

$$\rho_{X,Y} = \frac{\text{COV}(X,Y)}{\sqrt{\text{VAR}(X)}\sqrt{\text{VAR}(Y)}} \quad (2.3)$$

onde  $\rho_{X,Y}$  é a correlação entre  $X$  e  $Y$ .

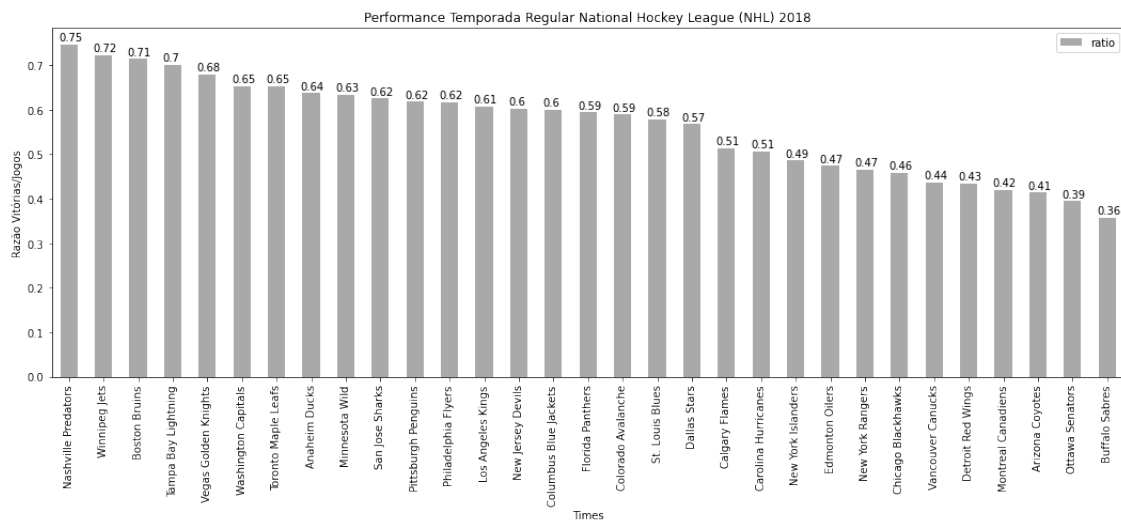
A matriz de correlação é uma ferramenta estatística capaz de medir a correlação de uma coleção de variáveis em seus pares, uma ótima maneira de obter de forma reduzida, o comportamento de várias séries em uma só tabela. Uma matriz de correlação de  $n$  variáveis  $X_1, \dots, X_n$ , é uma matriz  $n \times n$ , no qual o  $i, j$ -ésimo elemento da matriz é a correlação entre  $X_i$  e  $X_j$ , isto é,  $\rho_{X_i, X_j}$ .

### 3 DADOS

Os dados das regiões metropolitanas dos Estados Unidos utilizados nesse estudo foram obtidos através de Wikipédia (2016), onde uma tabela pode ser encontrada relacionando cada região metropolitana com os times da big four. Nesta tabela também há informações sobre a quantidade populacional de cada região, com dados atualizados para o ano de 2016.

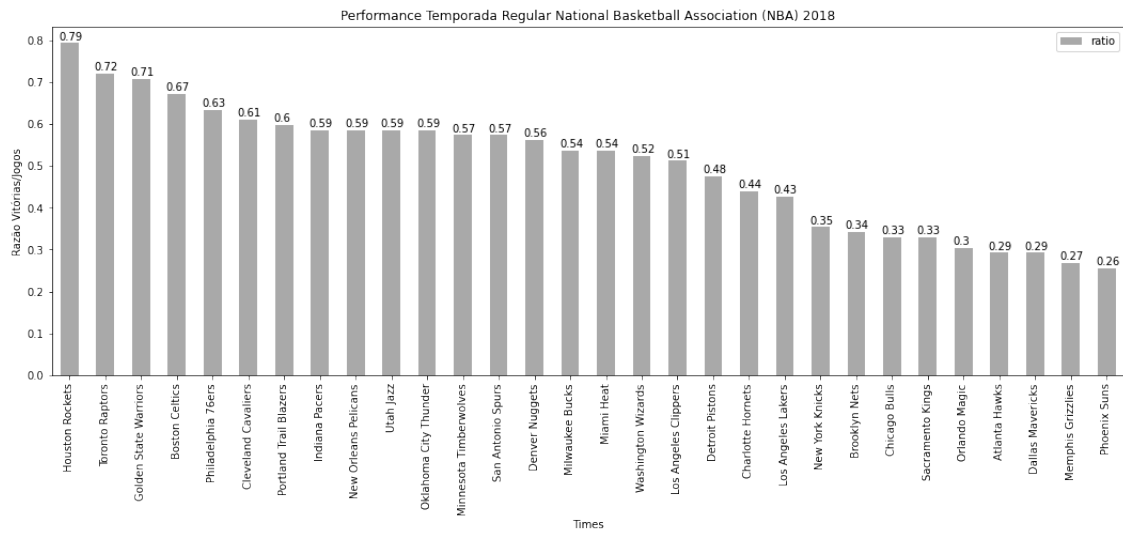
A primeira liga abordada foi a Liga Nacional de Hóquei. Para esta e todas as demais ligas presentes neste trabalho, foi escolhido os resultados da temporada regular de 2018. No gráfico 1, podemos verificar a performance de cada time na temporada regular utilizando o medidor proposto como a razão de vitórias e jogos.

**Figura 1** – Performance Temporada Regular National Hockey League (NHL) 2018

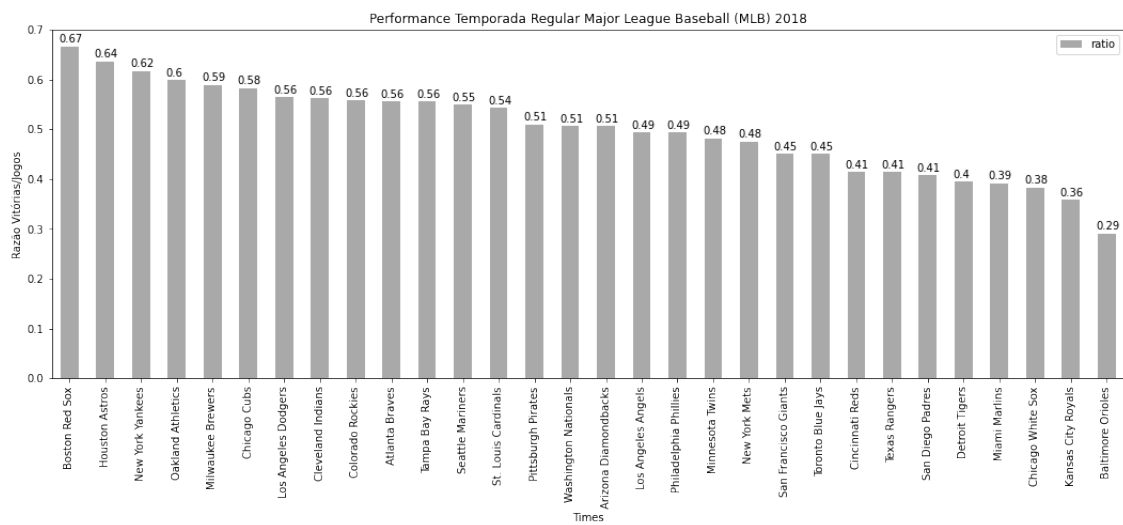


Fonte: Elaboração própria a partir de Reference (2018d)

O destaque para a liga de hóquei ficou para o Nashville Predators. O time de Nashville se mostrou muito forte com uma razão de 75% de performance. O segundo esporte analisado é o basquete,

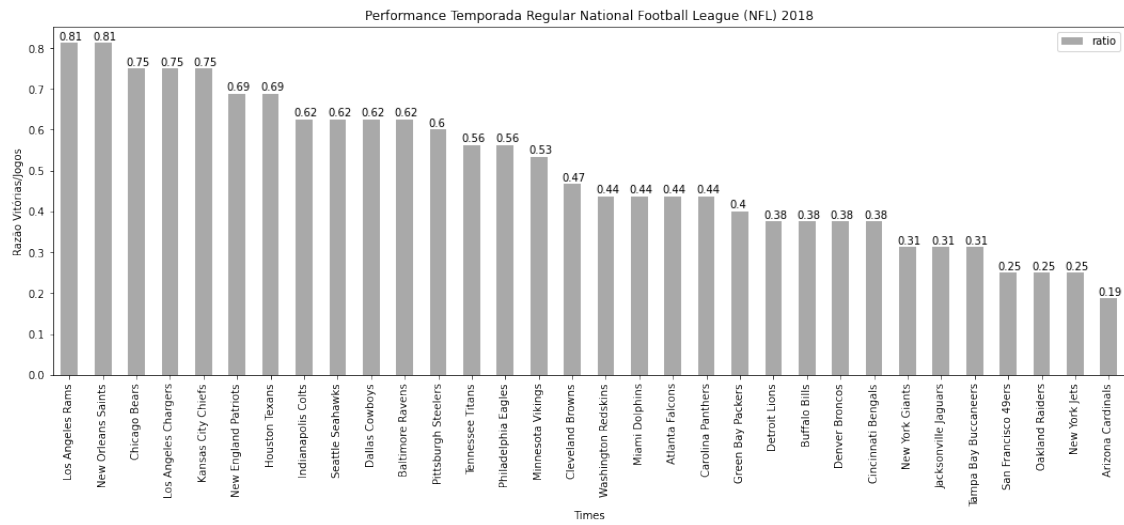
**Figura 2 – Performance Temporada Regular National Basketball Association (NBA) 2018**

Fonte: Elaboração própria a partir de Reference (2018b)

**Figura 3 – Performance Temporada Regular Major League Baseball (MLB) 2018**

Fonte: Elaboração própria a partir de Reference (2018a)

**Figura 4 – DPerformance Temporada Regular National Football League (NFL) 2018**



Fonte: Elaboração própria a partir de Reference (2018c)



## 4 CONCLUSÃO

	NFL	NBA	NHL	MLB
NFL		0.237000	0.304000	-0.050000
NBA	0.237000		0.760000	0.325000
NHL	0.304000	0.760000		0.432000
MLB	-0.050000	0.325000	0.432000	

## REFERÊNCIAS

REFERENCE. *Major League Baseball (MLB)*. 2018. Disponível em: <<https://www.baseball-reference.com/leagues/majors/2018.shtml>>. Acesso em: 25 fev. 2022. 7

REFERENCE. *National Basketball Association (NBA)*. 2018. Disponível em: <[https://www.basketball-reference.com/leagues/NBA\\_2018.html](https://www.basketball-reference.com/leagues/NBA_2018.html)>. Acesso em: 25 fev. 2022. 7

REFERENCE. *National Football League (NFL)*. 2018. Disponível em: <<https://www.pro-football-reference.com/years/2018/>>. Acesso em: 25 fev. 2022. 8

REFERENCE. *National Hockey League (NHL)*. 2018. Disponível em: <[https://www.hockey-reference.com/leagues/NHL\\_2018.html](https://www.hockey-reference.com/leagues/NHL_2018.html)>. Acesso em: 25 fev. 2022. 6

WIKIPÉDIA. *Regiões Metropolitanas (EUA)*. 2016. Disponível em: <[https://en.wikipedia.org/wiki/List\\_of\\_American\\_and\\_Canadian\\_cities\\_by\\_number\\_of\\_major\\_professional\\_sports\\_franchises](https://en.wikipedia.org/wiki/List_of_American_and_Canadian_cities_by_number_of_major_professional_sports_franchises)>. Acesso em: 25 fev. 2022. 6

WIKIPÉDIA. *Correlação*. 2021. Disponível em: <<https://pt.wikipedia.org/wiki/Correla%C3%A7%C3%A3o>>. Acesso em: 25 fev. 2022. 4

WIKIPÉDIA. *Correlação de Pearson*. 2021. Disponível em: <[https://pt.wikipedia.org/wiki/Coefficiente\\_de\\_correla%C3%A7%C3%A3o\\_de\\_Pearson](https://pt.wikipedia.org/wiki/Coefficiente_de_correla%C3%A7%C3%A3o_de_Pearson)>. Acesso em: 25 fev. 2022. 4