

# Olimpíadas Rio 2016

Bruno Vieira e Isabela Meneguci

# Introdução

- Existe uma relação entre o PIB de um país e a quantidade de medalhas ganhas?
- Países com maior PIB tendem a se sair melhor em esportes individuais?
- A idade de um atleta influencia na performance em esportes de alto rendimento?

# Metodologia

- 1ª pergunta: Regressão Linear
- 2ª pergunta: Teste de Hipótese
- 3ª pergunta: Algoritmo de Classificação- kNN

# Caracterização Inicial dos Dados

A base de dados utilizada foi retirada do Kaggle.

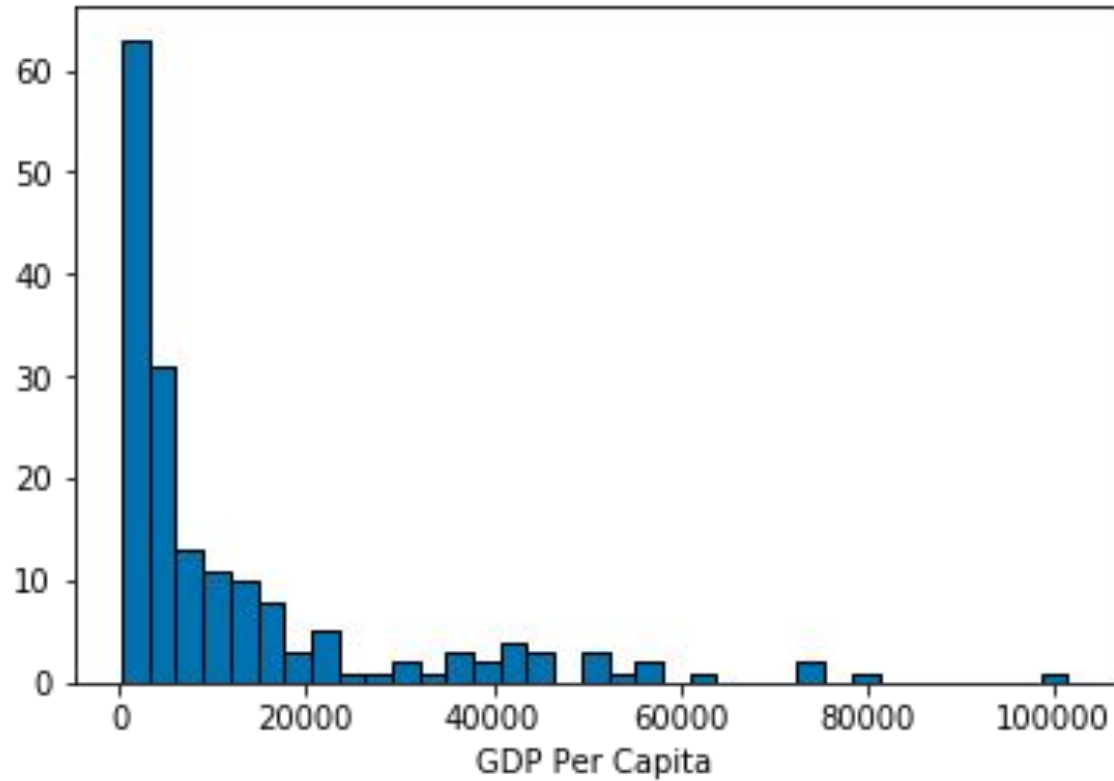
Ela contém 3 tabelas: athletes, countries e events.

Fizemos uma tabela auxiliar chamada sports.

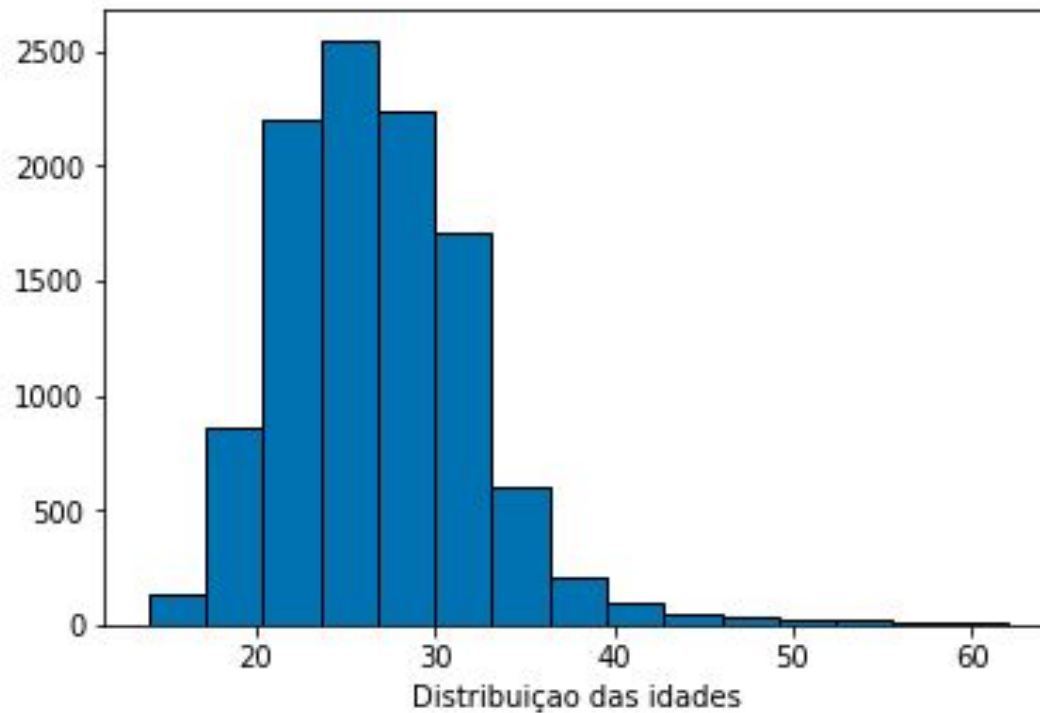
Criamos um data frame a partir dessas 4 tabelas. Adicionando a ele a idade do atleta.

	nationality	sex	dob	sport	gold	silver	bronze	single	group	country	population	gdp_per_capita	age	medals_sum
0	ESP	male	10/17/69	athletics	0	0	0	1	1	Spain	46418269.0	25831.582305	47.0	0
1	KOR	female	9/23/86	fencing	0	0	0	1	0	Korea, South	50617045.0	27221.524051	30.0	0
2	CAN	male	5/27/92	athletics	0	0	1	1	1	Canada	35851774.0	43248.529909	24.0	1
3	MDA	male	1/2/91	taekwondo	0	0	0	1	0	Moldova	3554150.0	1848.061804	25.0	0
4	NZL	male	11/26/90	cycling	0	0	0	1	0	New Zealand	4595700.0	37807.967276	26.0	0
5	AUS	male	1/26/90	triathlon	0	0	0	1	0	Australia	23781169.0	56310.962993	26.0	0
6	USA	male	6/4/93	volleyball	0	0	1	0	1	United States	321418820.0	56115.718426	23.0	1
7	AUS	male	9/25/91	aquatics	0	0	0	1	1	Australia	23781169.0	56310.962993	25.0	0
8	ESP	female	12/14/88	athletics	0	0	0	1	1	Spain	46418269.0	25831.582305	28.0	0
9	ETH	female	7/22/91	athletics	0	0	0	1	1	Ethiopia	99390750.0	619.169406	25.0	0
10	ETH	male	11/6/97	athletics	0	0	0	1	1	Ethiopia	99390750.0	619.169406	19.0	0
11	BRN	male	5/17/96	athletics	0	0	0	1	1	Bahrain	1377237.0	22600.214098	20.0	0
13	USA	female	5/25/92	athletics	0	0	0	1	1	United States	321418820.0	56115.718426	24.0	0
14	USA	female	12/3/96	aquatics	1	1	0	1	1	United States	321418820.0	56115.718426	20.0	2
15	GBR	female	4/10/96	rugby sevens	0	0	0	0	1	United Kingdom	65138232.0	43875.969614	20.0	0
16	UZB	male	7/7/98	wrestling	0	0	0	1	0	Uzbekistan	31299500.0	2132.070368	18.0	0
17	RSA	male	2/18/94	football	0	0	0	0	1	South Africa	54956920.0	5723.973357	22.0	0
18	NZL	female	11/20/89	football	0	0	0	0	1	New Zealand	4595700.0	37807.967276	27.0	0
19	EGY	male	6/3/89	volleyball	0	0	0	0	1	Egypt	91508084.0	3614.746766	27.0	0
20	MAR	male	3/25/87	athletics	0	0	0	1	1	Morocco	34377511.0	2878.201342	29.0	0

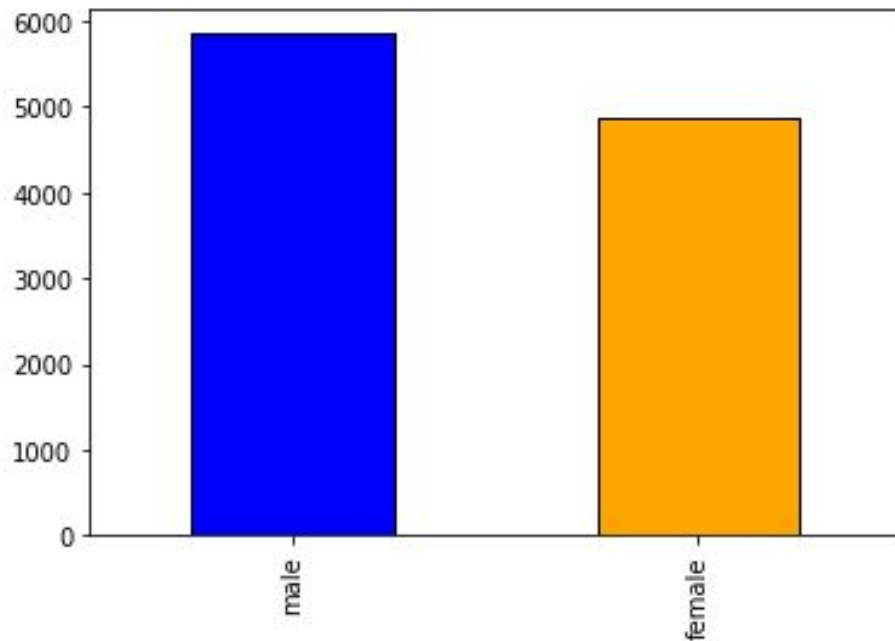
# GDP Per Capita



# Distribuição das idades



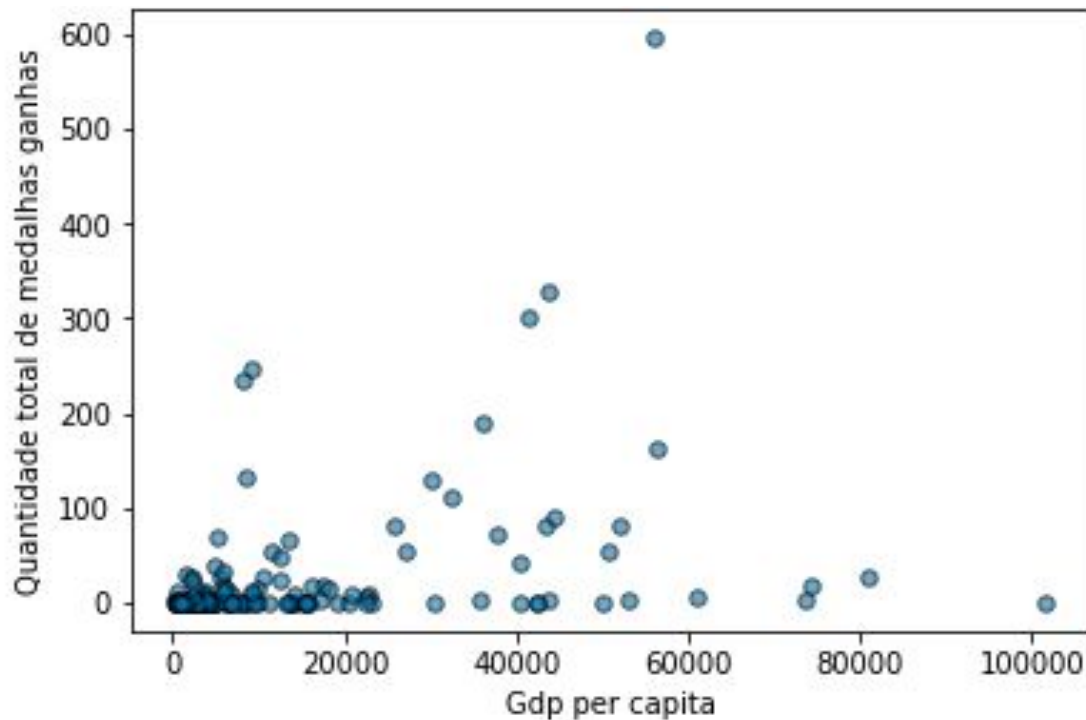
# Quantidade de atletas de cada sexo




Atletas do sexo feminino: 4866  
Atletas do sexo masculino: 5855




# Quantidade de medalhas X GDP per capita



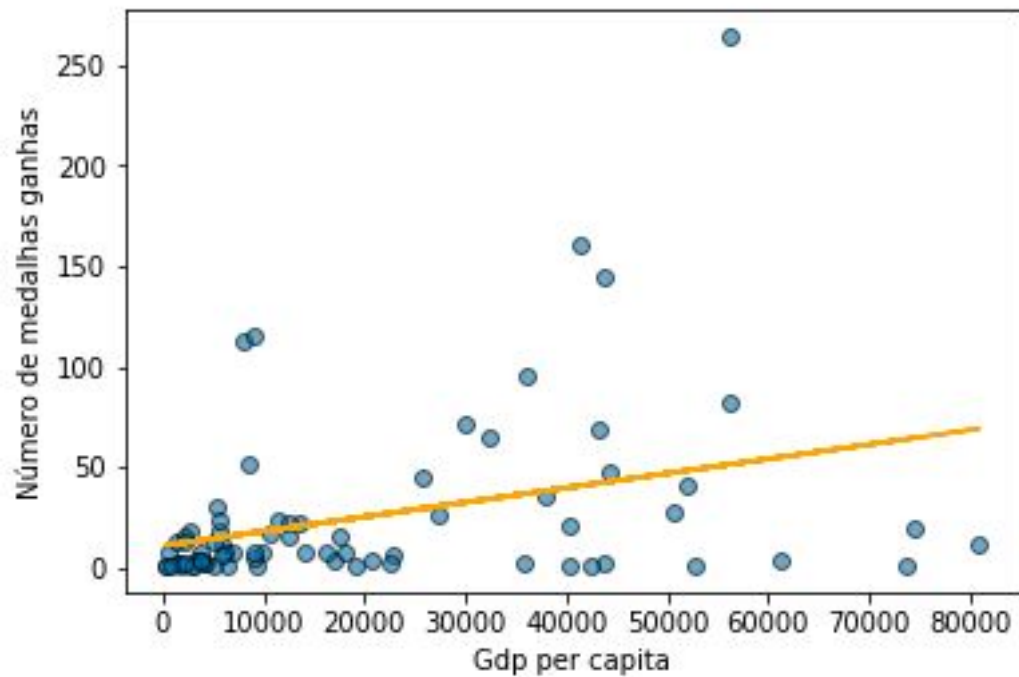
# Resultados e Conclusões

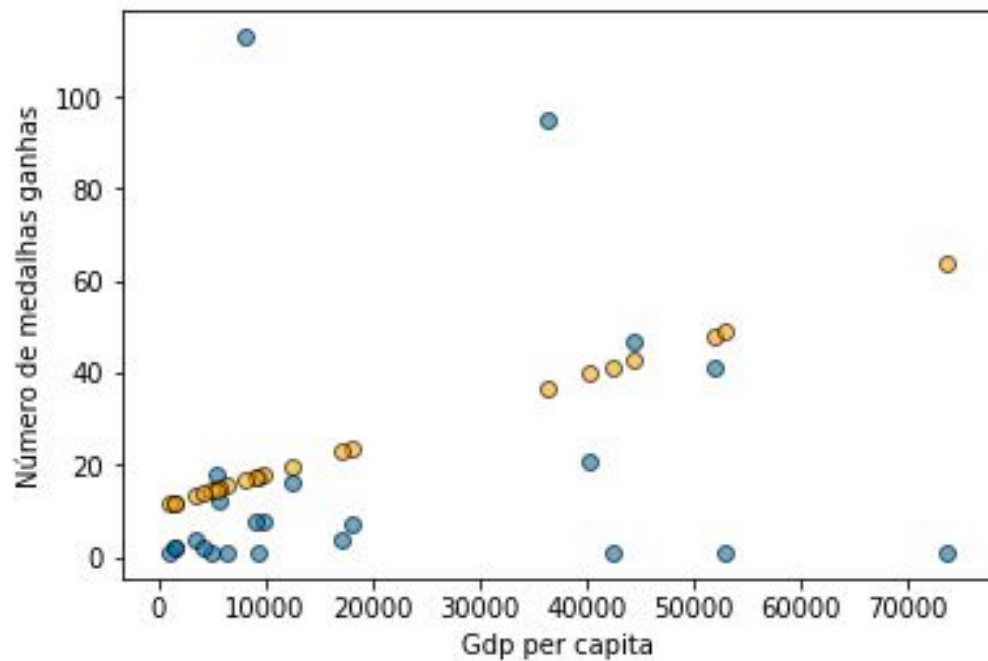


Existe uma relação entre o  
PIB de um país e a  
quantidade de medalhas  
ganhas?



# Regressão Linear Simples






Erro quadrado médio: 1167.0977807569127

Métrica R2: 0.10049082265770248


## Conclusão

Não conseguimos ajustar os dados de uma boa maneira :(

Mas podemos perceber pela reta da regressão e pelo fator  $R^2$  que existe uma relação crescente fraca entre o PIB e o número de medalhas ganhas, mas nada muito significativo.



Países com maior PIB  
tendem a se sair melhor em  
esportes individuais?



# Teste de Hipótese 1

- $H_0$  (hipótese nula): Um PIB elevado leva a um melhor desempenho **em pelo menos** um dos esportes individuais mais de 50% das vezes.
- $H_1$  (hipótese alternativa): Não é possível inferir uma relação entre PIB e desempenho em esportes individuais.



# Pais Melhor\_que\_a\_media

0	Australia	1
1	Austria	0
2	Canada	1
3	Denmark	1
4	Finland	0
5	Germany	1
6	Hong Kong*	0
7	Iceland	0
8	Ireland	1
9	Luxembourg	0
10	Netherlands	1
11	Norway	1
12	Qatar	0
13	Singapore	0
14	Sweden	1

# Bootstrap

Executamos um bootstrap com 10000 repetições para calcular a probabilidade de um país, selecionado ao acaso, ser melhor em pelo menos um dos esportes individuais que ele participa.

E...

O resultado gira em torno de 60%

Aceitamos a hipótese nula! (:

## Teste de Hipótese 2

- $H_0$  (hipótese nula): Um PIB elevado leva a um melhor desempenho em **pelo menos metade** dos esportes individuais que o país participa mais de 50% das vezes.
- $H_1$  (hipótese alternativa): Não é possível inferir uma relação entre PIB e desempenho em esportes individuais.

	Pais	Melhor_que_a_media
0	Australia	1
1	Austria	0
2	Canada	0
3	Denmark	1
4	Finland	0
5	Germany	1
6	Hong Kong*	0
7	Iceland	0
8	Ireland	0
9	Luxembourg	0
10	Netherlands	0
11	Norway	0
12	Qatar	0

# Bootstrap

Da mesma forma como foi feito anteriormente, fizemos um bootstrap com 10000 repetições. Mas agora para calcular a probabilidade de um país, selecionado ao acaso, ser melhor em pelo menos metade dos esportes individuais que ele participa.

E...

O resultado gira em torno de 27%


Rejeitamos a hipótese nula! :(

## Conclusão


- Pelo primeiro teste de hipótese: pode existir uma relação entre o país e esportes específicos incentivados.
- Pelo segundo teste de hipótese: vimos que dificilmente os países com PIB mais alto são bons em vários esportes individuais ao mesmo tempo.

No geral, podemos dizer que sim, países com maior PIB costumam se sair melhor em esportes individuais.

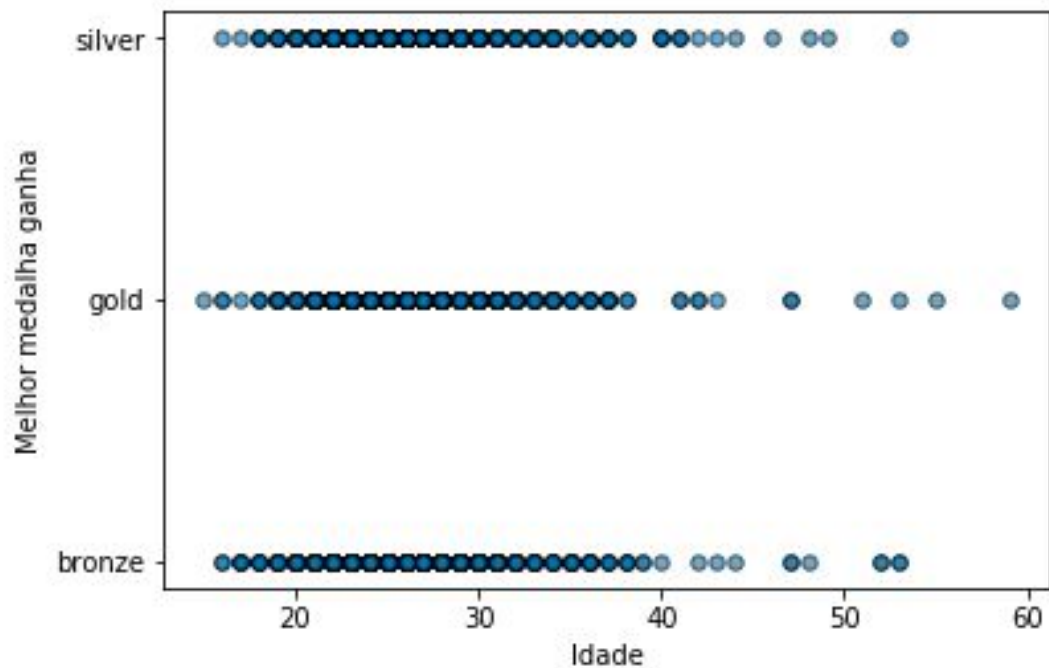


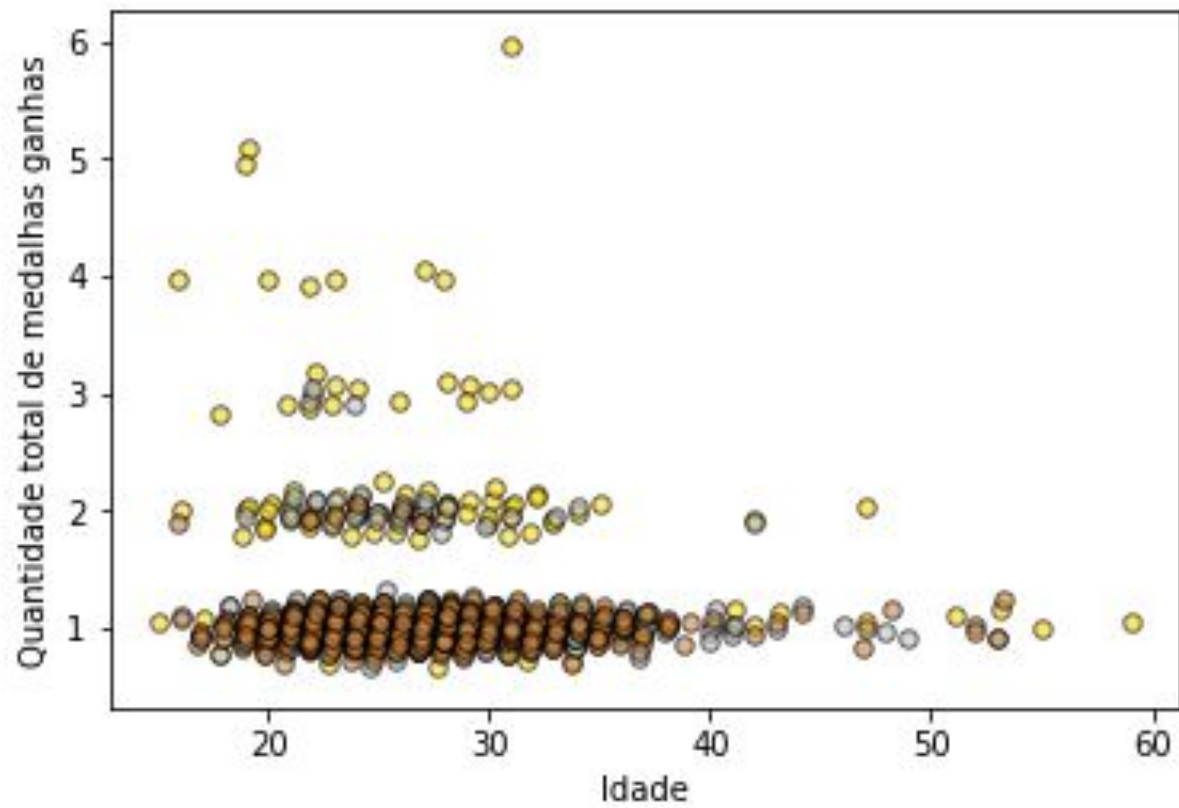


A idade de um atleta  
influencia na performance  
em esportes de alto  
rendimento?

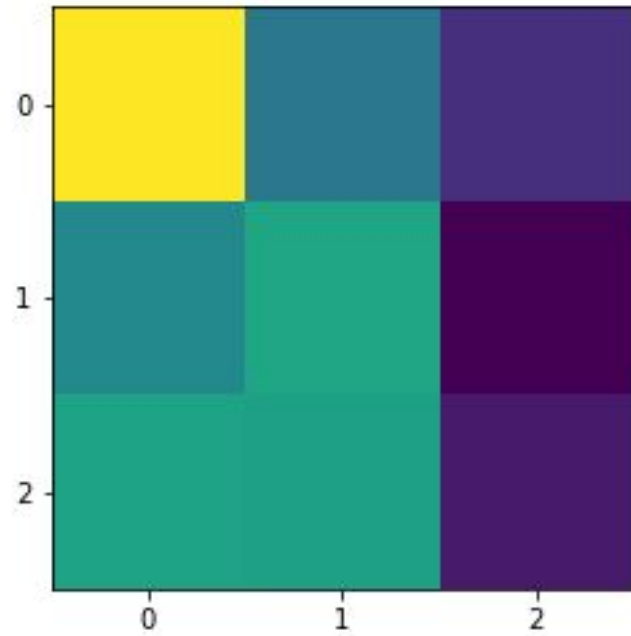


# k-Nearest Neighbor (kNN)





	precision	recall	f1-score	support
bronze	0.44	0.54	0.48	196
gold	0.36	0.46	0.40	158
silver	0.34	0.18	0.23	171
accuracy			0.39	525
macro avg	0.38	0.39	0.37	525
weighted avg	0.38	0.39	0.38	525



# Conclusão

Não podemos afirmar essa pergunta! :(