

# Análise Estatística do Desempenho de Atletas Olímpicos

Luciano Rodrigues

Ciência da Computação - UNESP

Junho de 2025

# Objetivo do Trabalho

- Aplicar estatística descritiva e probabilidade
- Investigar o perfil e desempenho de atletas olímpicos
- Prever a probabilidade de um atleta conquistar medalha

# Conjunto de Dados

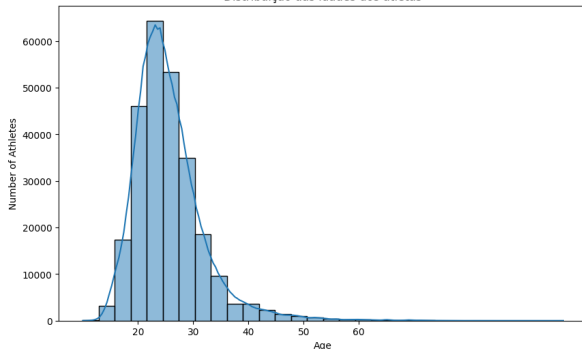
- 271.116 registros (1896–2016)
- Fonte: Kaggle *“120 Years of Olympic History”*
- Variáveis principais: idade, peso, altura, sexo, esporte, medalha

- Remoção de valores nulos
- Conversão de medalha para binária (1: medalhista, 0: não)
- Codificação de variáveis categóricas (sexo, esporte)
- Seleção de variáveis relevantes

# Análise Exploratória: Perfil Geral

- Idade média: 25 anos
- Peso médio: 70 kg
- Altura média: 175 cm
- Distribuições aproximam-se de normal ou gama

Distribuição das idades dos atletas



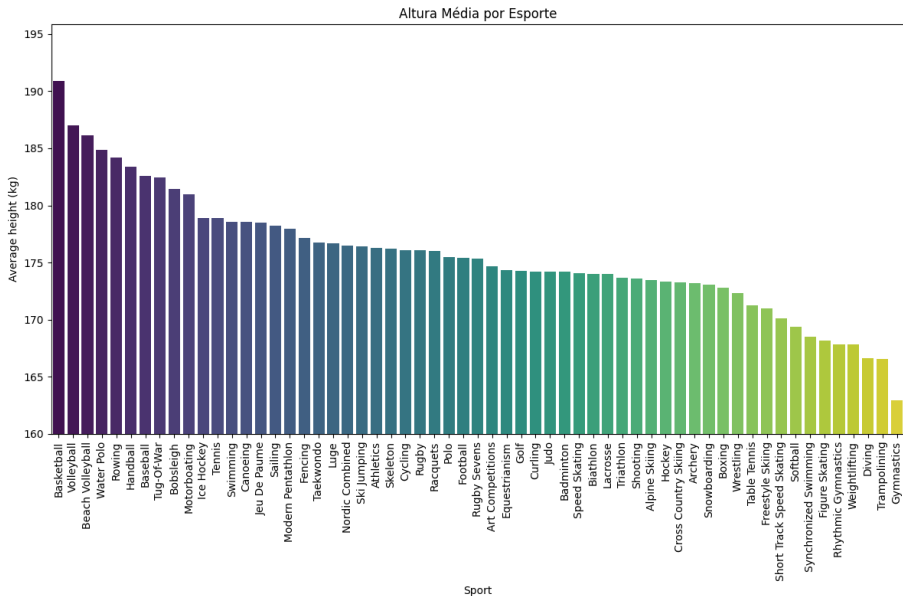
# Estatísticas Descritivas: Idade, Peso e Altura

Estatística	Idade	Peso (kg)	Altura (cm)
Total de registros válidos	261.642	208.241	210.945
Média	25,56	70,70	175,34
Desvio padrão	6,39	14,35	10,52
Mínimo	10	25	127
1º Quartil (25%)	21	60	168
Mediana (50%)	24	70	175
3º Quartil (75%)	28	79	183
Máximo	97	214	226

# Análise Exploratória: Altura Por Esporte

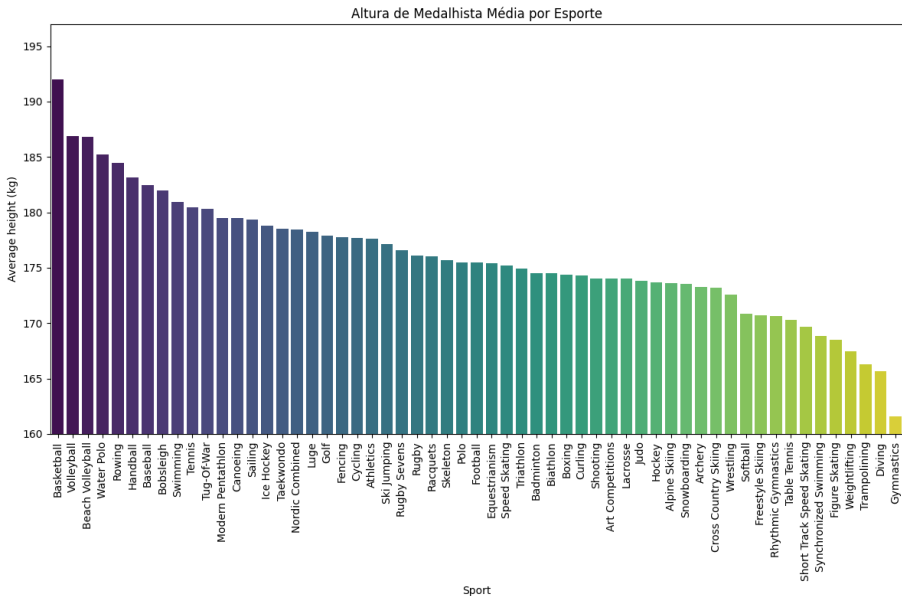
- Basquete: maior altura média (190 cm)
- Ginástica: menor altura média (161 cm)
- Diferenças refletem exigências físicas das modalidades

# Análise Exploratória: Altura Por Esporte



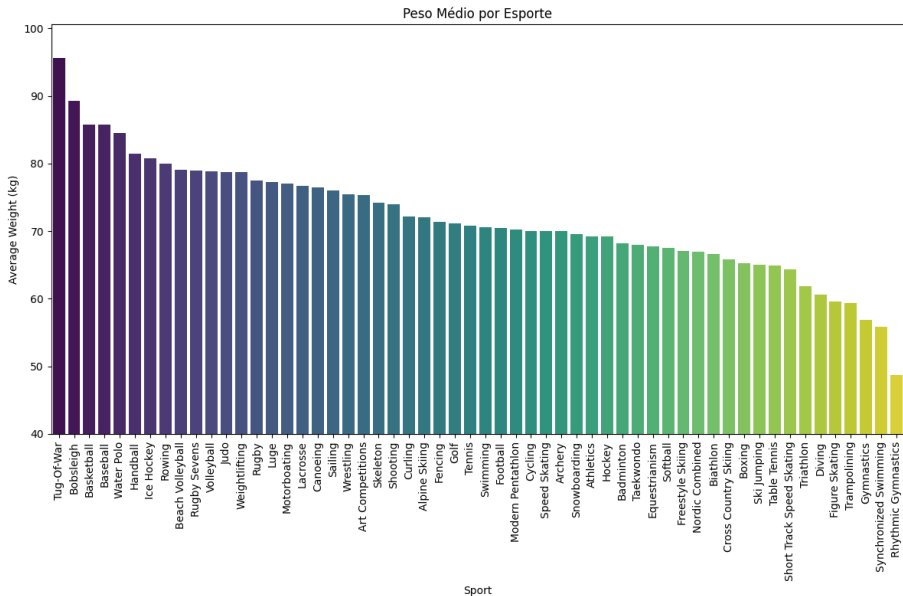


# Análise Exploratória: Altura Por Esporte

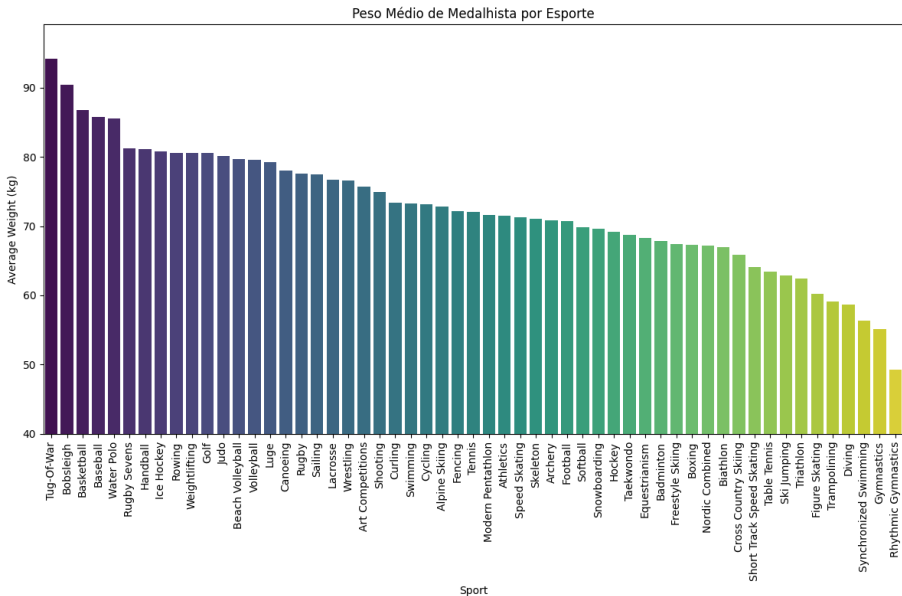


- Tug-of-War: maior peso médio (96 kg)
- Ginástica Rítmica: menor peso médio (48 kg)
- Perfis corporais variam de acordo com exigências das modalidades

# Análise Exploratória: Peso Por Esporte

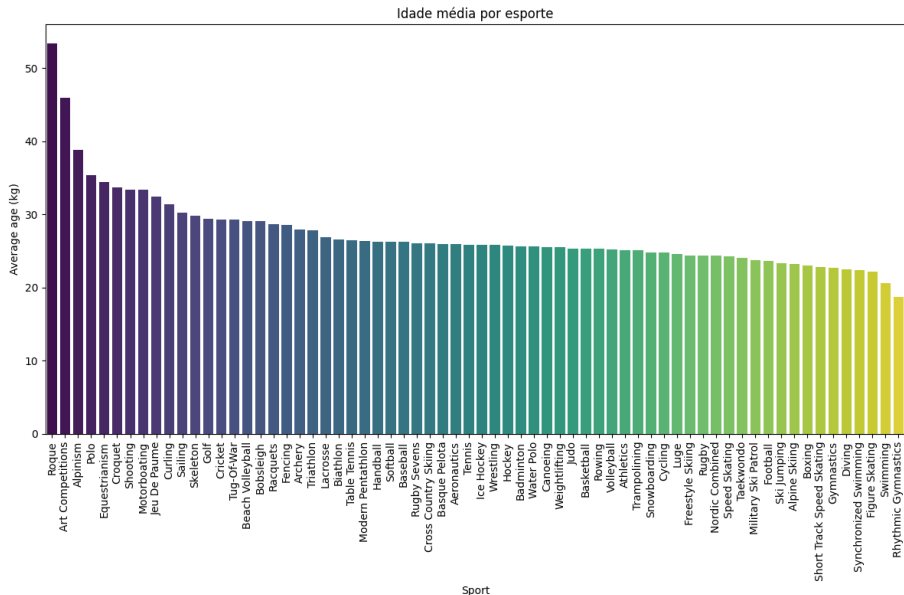


# Análise Exploratória: Peso Por Esporte



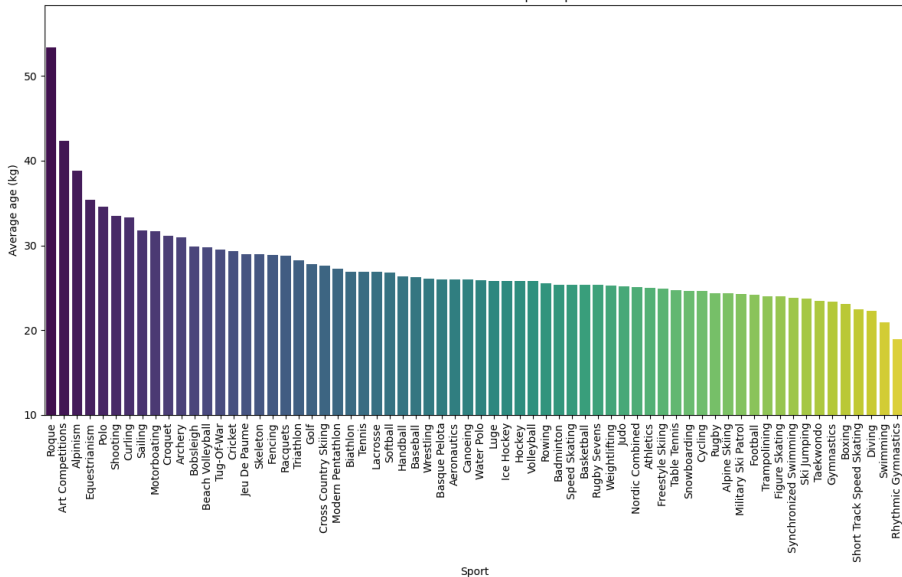
- Roque: maior idade média (50+ anos)
- Ginástica Rítmica: menor idade média (<20 anos)
- Modalidades técnicas favorecem atletas mais experientes ou jovens

# Análise Exploratória: Idade Por Esporte



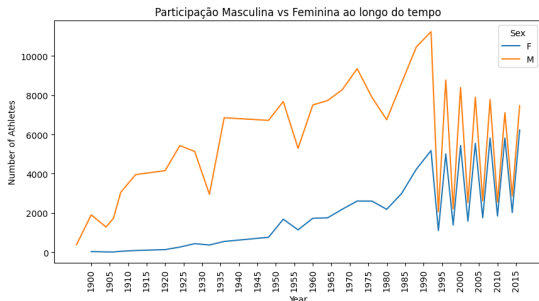
# Análise Exploratória: Idade Por Esporte

Idade de Medalhista Média por Esporte



# Análise Exploratória: Gênero

- Masculino: 196.594 atletas
- Feminino: 74.522 atletas
- Crescimento da participação feminina após 1994

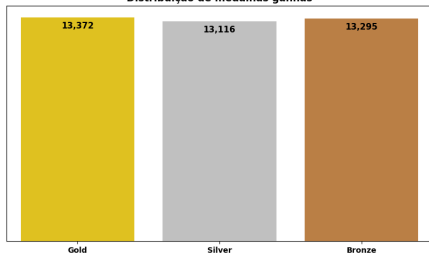




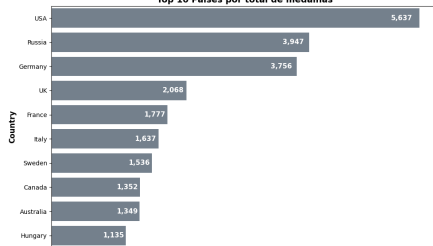
# Análise Exploratória: Medalhas

- Ouro → Bronze → Prata (leve diferença)
- Top 3: EUA, Rússia, Alemanha

Distribuição de medalhas ganhas



Top 10 Países por total de medalhas



# Predição: Preparação dos Dados

- Variáveis: Sexo, Idade, Peso, Altura, Esporte
- Alvo: Medalhista (1) ou Não (0)
- Modelos: Naive Bayes Gaussiano e Random Forest

- Acurácia: 85%
- Baixo *recall* para medalhistas (2%)
- Resultado próximo ao acaso

AUC-ROC: 0,59

## Classification Report:

Classe	Precisão	Revoc.	F1-score	Suporte
0	0.85	0.99	0.92	35190
1	0.28	0.02	0.03	6043
Acurácia			0.85	(41233 amostras)
Macro média	0.57	0.50	0.48	(41233 amostras)
Média ponderada	0.77	0.85	0.79	(41233 amostras)

# Random Forest (n=1000)

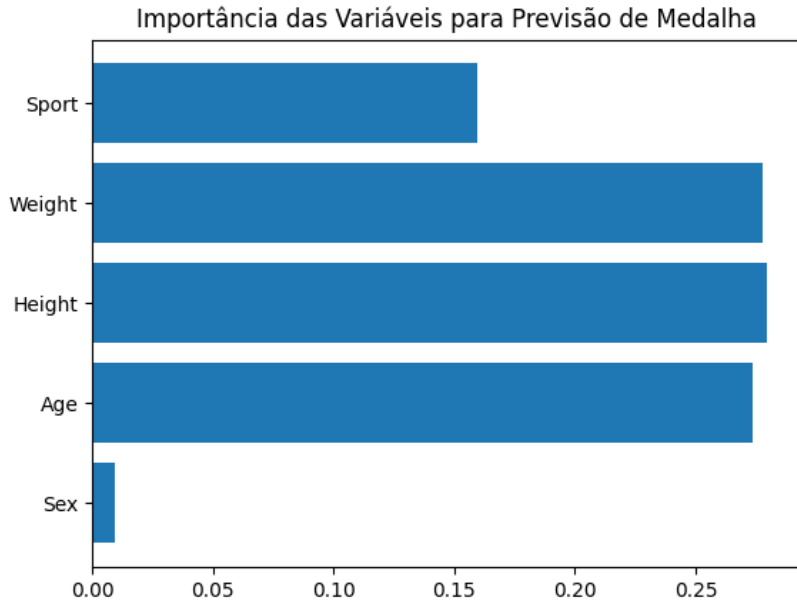
- Acurácia: 83%
- Melhor desempenho geral
- *Recall* para medalhistas: 17%

AUC-ROC: 0,68

## Classification Report:

Classe	Precisão	Revoc.	F1-score	Suporte
0	0.87	0.94	0.90	35190
1	0.34	0.17	0.23	6043
Acurácia			0.83	(41233 amostras)
Macro média	0.60	0.56	0.57	(41233 amostras)
Média ponderada	0.79	0.83	0.80	(41233 amostras)

# Random Forest (n=1000)



# Interpretação dos Resultados

- Diferenças corporais entre medalhistas e não medalhistas são pequenas
- Características como histórico, treino e apoio não estão nos dados
- Modelos têm desempenho limitado com informações físicas apenas

# Considerações Finais

- Estatística descritiva revela padrões históricos valiosos
- Predição limitada por falta de variáveis contextuais
- Trabalhos futuros: novos dados, modelos avançados, features externas

Obrigado pela atenção!

Perguntas?