

Análise Exploratória:

Total de medalhas nas Olimpíadas de Tóquio (2021)

Um projeto de portfolio por Júlio Zampietro

Principais conclusões

- Com relação ao total de medalhas por país, há forte correlação com PIB (0,831) e total de atletas competindo por país (0,875). Dados europeus também demonstram forte correlação com investimentos em esportes (0,817).
- Medidas per capita de PIB e investimentos em esportes têm fraca correlação com total de medalhas, o que indica que investimento por atleta é mais significativo do que por cidadão. Isso confere certa vantagem a países populosos.
- Países com quantidade de atletas em esportes coletivos acima da média (ver documentação) tenderam a se sair melhor. Isso pode parecer contraintuitivo, já que a razão oportunidades de medalha pelo total de atletas é menor nesses esportes. No entanto, a correlação com investimento total dá base para o argumento de que, ainda que esportes individuais sejam menos custosos para cada país competidor, aqueles que podem também investir em esportes coletivos tendem a se sobressair em termos de total de medalhas.
- Se recomenda o aumento no investimento em esportes para ambas as categorias para aumentar a quantidade de atletas participantes e, em consequência, o total de medalhas.

Construindo o *dataset* e matrizes de correlação

Dados foram coletados de várias fontes para construir um *dataset* inicial com dados a nível de país para total de medalhas, eficiência (ver documentação), total de atletas, PIB, PIB per capita, investimentos totais e per capita com educação, e índice GINI. A matriz de correlação (números absolutos) abaixo foi criada:

	Total_medals	Efficiency	Total_Athletes	Total_GDP	GDP_per_capita	Total_edu_spending	Per_capita_edu	Gini_coefficient
Total_medals	1	0,3873	0,8751	0,8314	0,2387	0,7705	0,1622	0,1873
Efficiency	0,3873	1	0,0857	0,2616	0,0271	0,2565	0,1888	0,1635
Total_Athletes	0,8751	0,0857	1	0,6678	0,2976	0,5467	0,2375	0,1078
Total_GDP	0,8314	0,2616	0,6678	1	0,1568	0,9914	0,019	0,0212
GDP_per_capita	0,2387	0,0271	0,2976	0,1568	1	0,0186	0,8681	0,3756
Total_edu_spending	0,7705	0,2565	0,5467	0,9914	0,0186	1	0,0227	0,051
Per_capita_edu	0,1622	0,1888	0,2375	0,019	0,8681	0,0227	1	0,2636
Gini_coefficient	0,1873	0,1635	0,1078	0,0212	0,3756	0,051	0,2636	1

Um segundo *dataset*, mais específico, foi criado a partir de dados europeus, visto que esta foi a única região quanto à qual foi simples encontrar dados de investimentos em esportes (total e per capita):

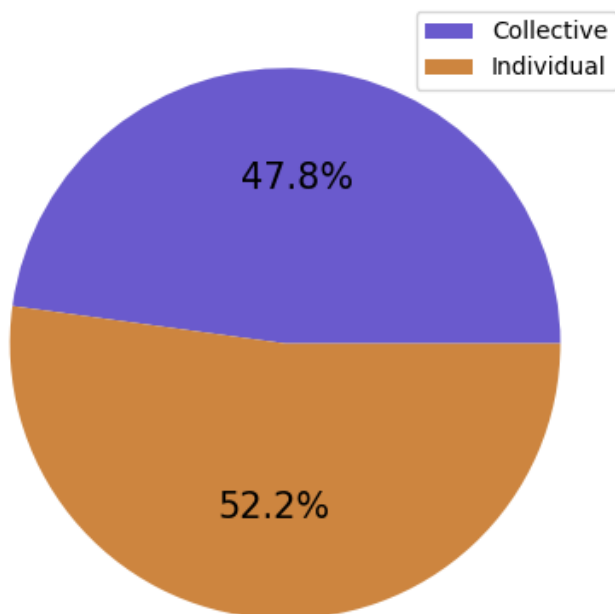
	Total_medals	Efficiency	Total_Athletes	Total_GDP	GDP_per_capita	Total_edu_spending	Per_capita_edu	Gini_coefficient	Total_sport_spending	Eurosport_per_capita
Total_medals	1	0,4355	0,9204	0,8263	0,0993	0,809	0,0138	0,016	0,8171	0,1962
Efficiency	0,4355	1	0,147	0,104	0,1346	0,1069	0,0312	0,1668	0,1137	0,0865
Total_Athletes	0,9204	0,147	1	0,9051	0,1331	0,8908	0,0111	0,0579	0,8997	0,1442
Total_GDP	0,8263	0,104	0,9051	1	0,1863	0,9866	0,0829	0,1056	0,9246	0,1482
GDP_per_capita	0,0993	0,1346	0,1331	0,1863	1	0,2699	0,9028	0,2475	0,1203	0,5005
Total_edu_spending	0,809	0,1069	0,8908	0,9866	0,2699	1	0,2027	0,0723	0,9241	0,2441
Per_capita_edu	0,0138	0,0312	0,0111	0,0829	0,9028	0,2027	1	0,1962	0,0682	0,6927
Gini_coefficient	0,016	0,1668	0,0579	0,1056	0,2475	0,0723	0,1962	1	0,0532	0,3007
Total_sport_spending	0,8171	0,1137	0,8997	0,9246	0,1203	0,9241	0,0682	0,0532	1	0,0187
Eurosport_per_capita	0,1962	0,0865	0,1442	0,1482	0,5005	0,2441	0,6927	0,3007	0,0187	1

- Total_medals tem forte correlação com Total_Athletes, Total_GDP, Total_edu_spending e Total_sport_spending, e não tem boa correlação com suas variantes per capita.
- Como há correlação quase perfeita (0,991) entre Total_edu_spending e Total_GDP, presumimos que o primeiro é função do segundo, e que Total_GDP é a métrica relevante.
- Total_sport_spending também se correlaciona fortemente com Total_Athletes (0,899).

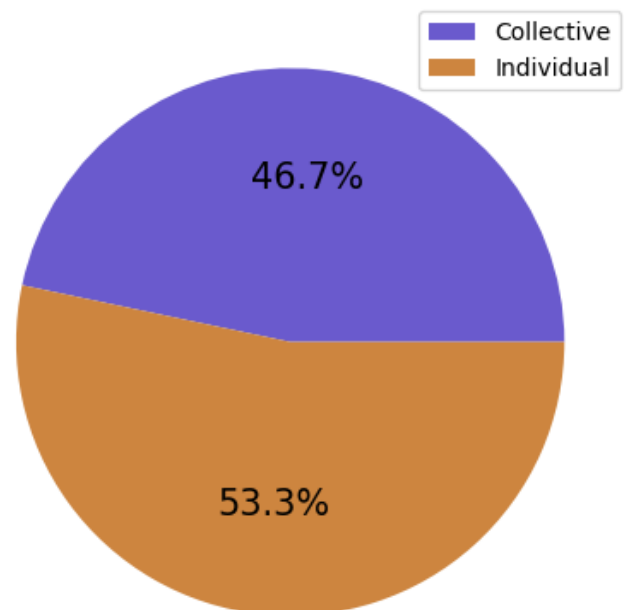
Esportes individuais vs. esportes coletivos

- SQL foi usado para obter o total de atletas competindo em modalidades coletivas (5174, 46,67%) e individuais (5911, 53,32%).
- SQL também foi usado para criar uma tabela de classificação de esportes, divididos em coletivos e individuais. Tanto atletas quanto esportes se dividem quase igualmente nas categorias:

Percentage of sports in each category

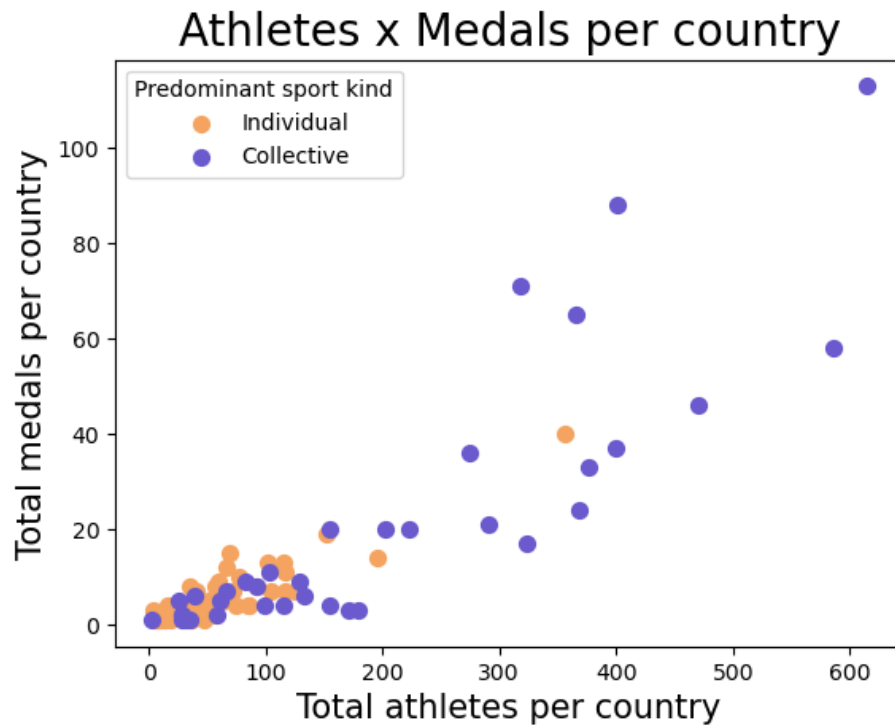


Percentage of athletes per category of sport

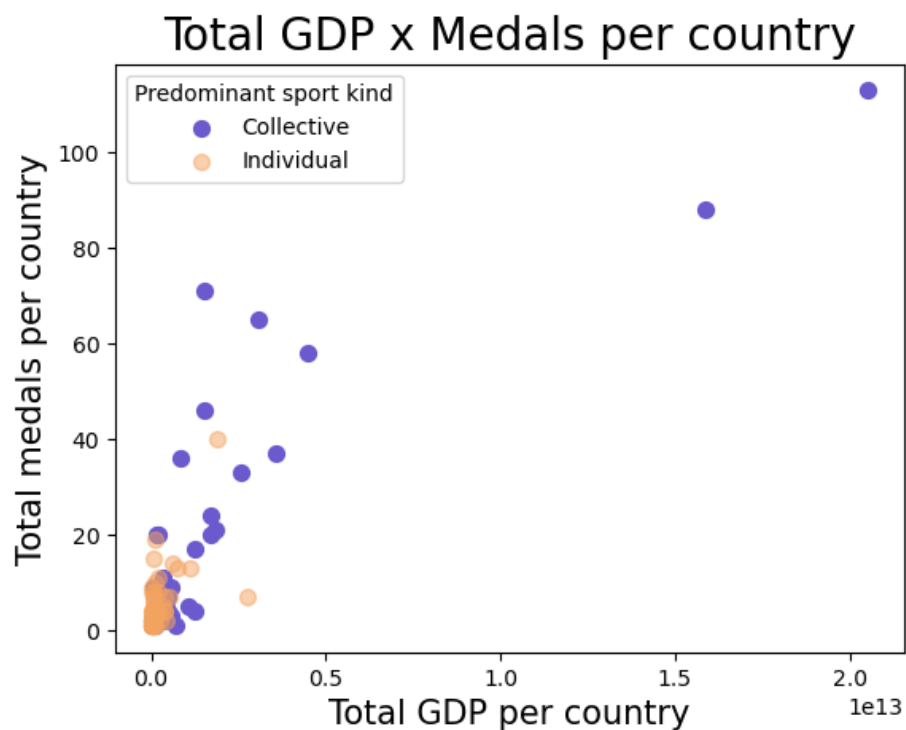


- Um claro padrão emerge entre as duas categorias quando as três variáveis mais relevantes são plotadas:

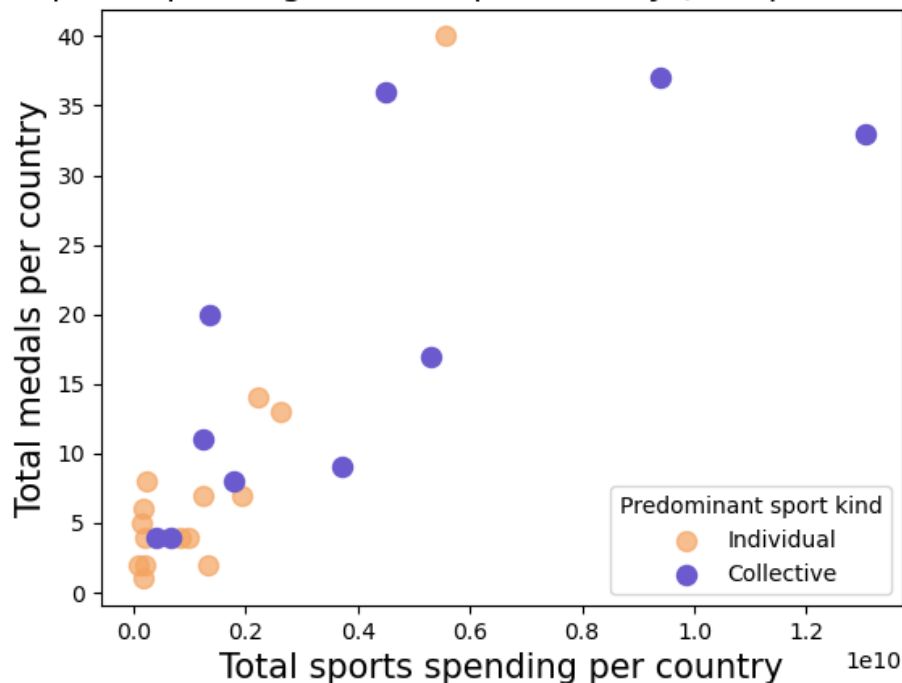
Graph 1



Graph 2



Graph 3 Sports spending x Medals per country (European Union)



- Enquanto países com mais atletas individuais que a média se mantêm nas piores colocações no ranking de medalhas totais, os com mais atletas coletivos tendem a se classificar melhor.
- O padrão não é explicado a partir da distribuição total de atletas nas categorias, já que esta é semelhante à distribuição de esportes. Também não é explicado pela eficiência, como comprovado pelo baixo índice de correlação.
- Uma explicação plausível é de que enquanto esportes individuais requerem menor investimento total para render medalhas, países com recursos econômicos para investir em ambas as categorias tendem a obter mais sucesso no todo.
- Concluimos que uma maneira de aumentar o total de medalhas é aumentar investimentos em esportes em ambas as categorias, o que levará a um aumento na participação de atletas (correlação de 0,899 com investimento em esportes) e, como consequência, um aumento no total de medalhas.