

Jogos Olímpicos à lupa

1st Duarte Meneses

Departamento de Engenharia Informática
Universidade de Coimbra
Coimbra, Portugal
duartemeneses@student.dei.uc.pt

2nd Patricia Costa

Departamento de Engenharia Informática
Universidade de Coimbra
Ovar, Portugal
patriciacosta@student.dei.uc.pt

Abstract—Os Jogos Olímpicos, também conhecidos como Olimpíadas, são considerados o maior evento desportivo a nível global. Essa importância leva a que a competição seja vista pelas nações como uma excelente oportunidade para demonstrarem o seu poderio e se destacarem em detrimento de outras. É com isso em vista que o nosso trabalho pretende ajudar cada país a identificar as modalidades em que não tem tão bons desempenhos e, assim, perceber onde pode direcionar o seu foco de desenvolvimento. Por outro lado, a nossa aplicação permitirá comparar o desempenho das nações em cada modalidade, o que pode ajudar os apostadores a perceber em quem devem apostar.

Index Terms—Jogos Olímpicos, Olimpíadas, modalidades, medalhas, nações

I. INTRODUÇÃO

Os Jogos Olímpicos [1] são o maior evento multidesportivo do mundo. Não é de admirar, por isso, que qualquer atleta ambiciona chegar ao tão mítico ouro olímpico.

A origem das Olimpíadas remonta à Grécia antiga onde representantes das várias cidades-estado competiam pela glória nas diversas modalidades. Mais recentemente, com a criação dos Jogos Olímpicos da era moderna, essa competição foi alargada às nações de todo o mundo e realiza-se de 4 em 4 anos.

Do ponto de vista desportivo, os Jogos Olímpicos são o ponto mais alto da carreira de um atleta. No entanto, têm também uma grande importância na componente política, visto serem uma excelente oportunidade para os diversos estados demonstrarem o seu poder. A histórica animosidade entre Estados Unidos e Rússia (ex URSS) sempre foi extrapolada nos Jogos Olímpicos. Ambas as nações fazem de tudo para se superiorizar de modo a passar uma imagem de força e grandiosidade ao mundo.

Embora seja a mais mediática, esta não é a única rivalidade política a passar para a competição das Olimpíadas. No entanto, mesmo sem estas animosidades, qualquer país ambiciona atingir o ouro uma vez que simboliza poder.

Fica assim evidente a importância de analisar onde cada nação pode evoluir nesta competição. É com isso em vista que pretendemos dotar os países de meios de análise dos seus resultados ao longo dos anos, através da visualização de informação. Deste modo, qualquer estado pode identificar em que modalidades não apresenta resultados tão positivos e, assim, direcionar a sua atenção para melhorar o seu desempenho nessas provas.

Por outro lado, ao identificar as modalidades em que uma nação não é tão boa, conseguimos também saber aquelas em que são melhores. Com isto, o nosso trabalho torna-se também útil aos apostadores: ao saberem que países se saem melhor em cada prova, conseguem ter uma maior confiança nas suas apostas.

II. TRABALHO RELACIONADO

Os Jogos Olímpicos, ao serem o maior evento desportivo à escala global, são imensamente analisados e estudados. Com isto, existem inúmeros datasets e trabalhos relacionados com este. Aquele em que nos inspiramos é da autoria de Marco Giuseppe de Pinto [2] que analisou diversas informações sobre as Olimpíadas, entre elas, a distribuição das medalhas de ouro pelos diversos países e eventos desportivos.

III. DADOS

Para desenvolvermos este trabalho, optamos por utilizar o conjunto de dados disponível no Kaggle intitulado de *Olympic Historical Dataset From Olympedia.org* [3]. Este é composto por 6 datasets, cada um com um propósito, interligando-se entre si de modo a complementar-se.

Consideramos que os datasets *Olympics_Country.csv* e *Olympic_Results.csv* contêm informação redundante e/ou desnecessária tendo em conta o problema em causa, pelo que apenas vamos utilizar os outros 4.

Nem todos os atributos se encontram listados abaixo pois decidimos apenas fazê-lo para aqueles que realmente nos interessam.

A coluna *NaNs* indica a quantidade de valores em falta em cada atributo. Em nenhum dos datasets a falta dessa informação prejudica a análise que pretendemos efetuar neste trabalho. Apenas no atributo *medal* do *Olympics_Athlete_Event_Results.csv* simboliza que o atleta não recebeu nenhum tipo de medalha. Neste caso, os valores *NaN* serão tratados adequadamente tendo em conta o contexto, uma vez que não significa desconhecimento do valor.

A. *Olympic_Athlete_Bio*

Cada entrada na tabela representa um atleta olímpico e a sua informação biológica. Tem os seguintes atributos:

Tabela I
INFORMAÇÕES SOBRE O DATASET *Olympic_Athlete_Bio.csv*

Nome	Descrição	Dtype	NaNs
athlete_id	Identificador único do atleta	int64	0
name	Nome completo do atleta	object	0
sex	Gênero do atleta	object	0
born	Data de nascimento do atleta	object	4284
height	Altura do atleta, em cm	object	50547
weight	Peso do atleta, em kg	object	50547
country	País que o atleta representa	object	0
country_noc	Abreviatura do país que o atleta representa	object	0
description	Breve descrição do atleta	object	103749
special_notes	Algumas notas sobre o atleta	object	94638

B. Olympic_Games_Medal_Tally

Cada entrada na tabela representa um país que tenha pelo menos uma medalha em cada um dos Jogos Olímpicos.

Tabela II
INFORMAÇÕES SOBRE O DATASET *Olympic_Games_Medal_Tally.csv*

Nome	Descrição	Dtype	NaNs
edition	Ano – Jogos Olímpicos de Verão/Inverno	object	0
edition_id	Identificador único da edição	int64	0
year	Ano em que as Olimpíadas ocorreram	int64	0
country	País em causa	object	0
country_noc	Sigla do país em causa	object	0
gold	Total de medalhas de ouro	int64	0
silver	Total de medalhas de prata	int64	0
bronze	Total de medalhas de bronze	int64	0
total	Total de medalhas	int64	0

C. Olympic_Athlete_Event_Results

Cada entrada na tabela contém um evento específico com os dados da participação de um atleta olímpico.

Tabela III
INFORMAÇÕES SOBRE O DATASET *Olympic_Athlete_Event_Results.csv*

Nome	Descrição	Dtype	NaNs
edition	Ano - Jogos Olímpicos de Verão/Inverno	object	0
edition_id	Identificador único da edição	int64	0
country_noc	Sigla do país que o atleta representa	object	0
sport	Desporto em questão	object	0
event	Evento específico do desporto em questão	object	0
result_id	Identificador único do evento	int64	0
athlete	Atleta que está a competir	object	0
athlete_id	Identificador único do atleta	int64	0
pos	Posição em que o atleta ficou na prova	object	0
medal	Se o atleta ganhou ou não uma medalha. Em caso afirmativo, se foi ouro, prata ou bronze	object	270336
isTeamSport	Indica se é ou não um evento de equipa	bool	0

D. Olympics_Games

Cada entrada na tabela representa uma edição dos Jogos Olímpicos.

Tabela IV
INFORMAÇÕES SOBRE O DATASET *Olympics_Games.csv*

Nome	Descrição	Dtype	NaNs
edition	Ano – Jogos Olímpicos de Verão/Inverno	object	0
edition_id	Identificador único da edição	int64	0
city	Cidade em que ocorreram os Jogos Olímpicos	object	0
country_noc	Sigla do país em que ocorreram as Olimpíadas	object	0

Analisando os dados, percebemos que existiam referências a nações que, atualmente, já não existem. São exemplo disso a União Soviética e a Jugoslávia. Nestes casos, é impossível atribuir as medalhas a um país em específico visto que estes se subdividiram em vários (por exemplo, a Jugoslávia dividiu-se em Bósnia, Croácia, Macedónia, Montenegro, Eslovénia e Sérvia). Aliada a esta situação, encontramos também referências a Equipa Mista, Atletas Olímpicos Individuais,... Em qualquer um destes casos, optamos por omitir os resultados.

Em sentido contrário, existem nações que se juntaram com o tempo. É o caso da Alemanha Federal e a Alemanha Democrática. Nestes casos, consideramos que as medalhas arrecadadas por estas duas nações pertenciam à Alemanha. Em suma, estas foram as nossas decisões:

Tabela V
DECISÕES RELATIVAMENTE À ATRIBUIÇÃO DE MEDALHAS AOS PAÍSES ATUAIS

Entidade	Decisão	Razão
Equipa Mista	Ignorar dados	Impossível atribuir as medalhas a um país em concreto
Equipa Unificada	Ignorar dados	Impossível atribuir as medalhas a um país em concreto
Atletas Olímpicos Individuais	Ignorar dados	Impossível atribuir as medalhas a um país em concreto
Bohemia	Ignorar dados	Dividiu-se em vários países
Australásia	Ignorar dados	Dividiu-se em vários países
Checoslováquia	Ignorar dados	Dividiu-se em vários países
Jugoslávia	Ignorar dados	Dividiu-se em vários países
União Soviética	Ignorar dados	Dividiu-se em vários países
República Árabe Unida	Ignorar dados	Dividiu-se em vários países
Federação das Índias Ocidentais	Ignorar dados	Dividiu-se em vários países
República Federal da Alemanha	Juntar dados aos da Alemanha	União de nações
República Democrática Alemã	Juntar dados aos da Alemanha	União de nações
Hong Kong	Juntar dados aos da China	Hong Kong é uma região da China
Comité Olímpico Russo	Juntar dados aos da Rússia	Os atletas são russos

IV. ANÁLISE EXPLORATÓRIA DOS DADOS

Com os dados à nossa disposição, conseguimos facilmente identificar que os Estados Unidos são o país com mais medalhas de cada tipo na generalidade dos eventos desportivos existentes.

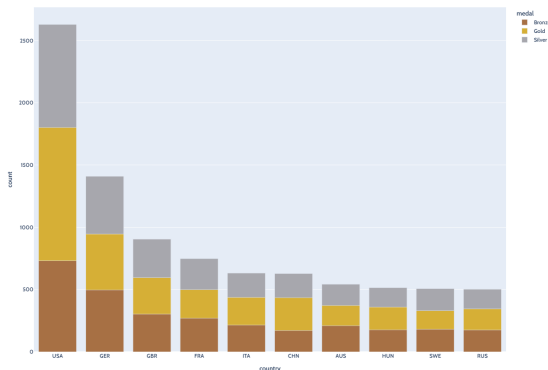


Figura 1. Heat map que representa o número de medalhas arrecadadas por cada nação ao longo de todas as edições das Olimpíadas

Analisando os países com mais medalhas no atletismo, percebemos que também são os Estados Unidos que dominam esse ranking.

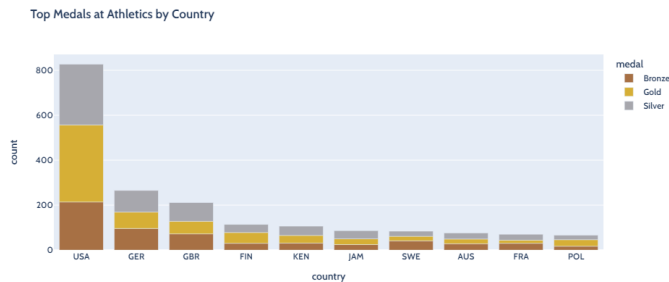


Figura 2. Top 10 de países com mais medalhas no Atletismo em todas as edições dos Jogos Olímpicos

V. REQUISITOS DE DESIGN

Nesta secção iremos mostrar como pretendemos que as visualizações da informação sejam e o modo como devem estar relacionadas. Todas as mockups aqui apresentadas podem ser consultadas com mais detalhe através da referência [12].

A. Vista principal

Pretendemos que a página inicial da nossa aplicação contenha um heat map do mapa mundo em que cada país apresenta a informação do número total de medalhas arrecadadas ao longo dos anos. Juntamente com este mapa, devem existir duas opções: uma para analisar os dados por nação e outro por modalidade. No geral, esta vista deve ter o seguinte aspeto (os gráficos mais à esquerda ainda não se encontram definidos):



Figura 3. Vista da análise geral (principal)

As visualizações que constam nesta vista são:

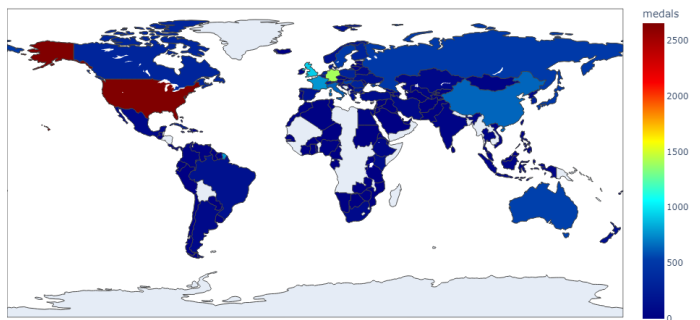
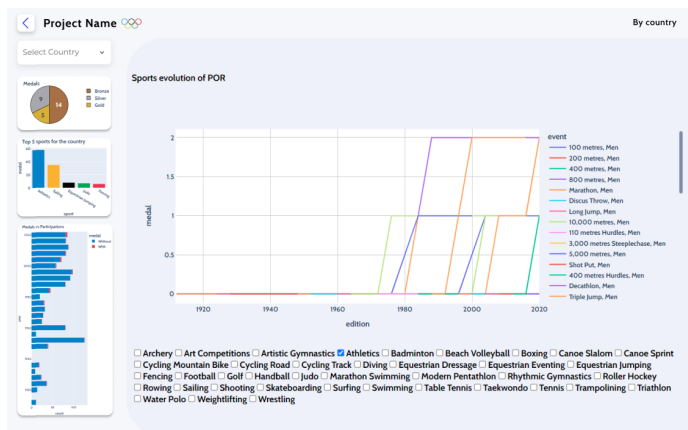


Figura 4. Heat map que representa o número de medalhas arrecadadas por cada nação ao longo de todas as edições das Olimpíada

Neste mapa temos alguns problemas devido à discrepância do número de medalhas entre os Estados Unidos e os restantes países. Isso leva a que as nações com menos medalhas fiquem com uma cor muito semelhante entre si, sendo quase impossível distinguí-las.

B. Vista das análises por país

Se optarmos por analisar os dados por país, deve ser possível visualizar a evolução do número de medalhas desse mesmo país em cada modalidade. Este gráfico deve ser um *line chart* com as modalidades selecionáveis. No lado esquerdo da página devem constar outras três visualizações: uma em que seja possível analisar o número e tipo de medalhas obtidas por essa nação (Figura 5), outra em que facilmente se identifiquem as cinco melhores modalidades desse estado (Figura 6) e ainda outra que explice a percentagem de medalhas tendo em conta as participações (Figura 7). Consideramos este último gráfico importante uma vez que, à partida, ganhar 3 medalhas é melhor do que 2. No entanto, se essas 2 medalhas forem arrecadadas em 5 participações, têm mais valor do que as 3 em 100 participações. Em suma, pretendemos que esta vista tenha a seguinte disposição:



Os gráficos presentes nesta vista são os seguintes:

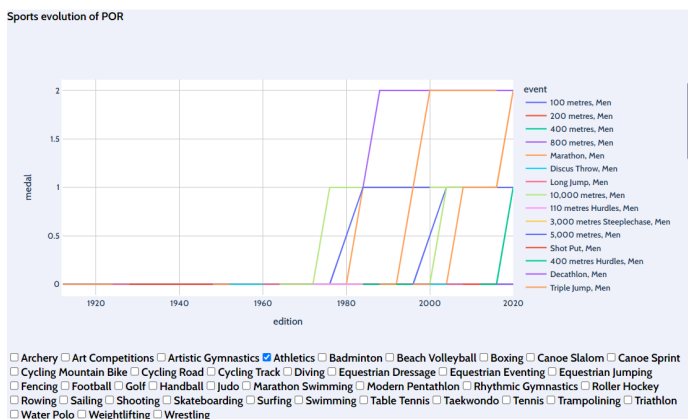


Figura 6. Evolução das medalhas conquistadas pela nação a analisar em cada modalidade

Na visualização acima encontramos claros problemas devido às inúmeras categorias de desportos e eventos. Pensamos resolver esta situação agrupando as modalidades. Isto não só resolveria o problema mencionado, como permitiria outras análises.

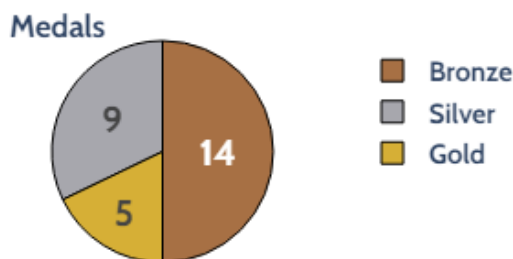


Figura 7. Tipos de medalhas arrecadas pela nação a analisar em todas as edições dos Jogos Olímpicos

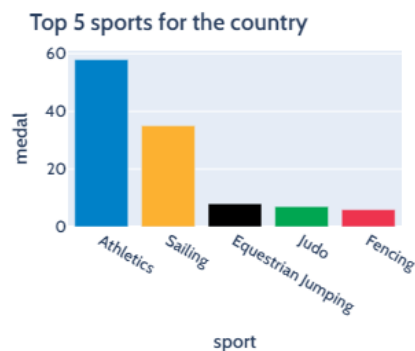


Figura 8. Top 5 de desportos do país a analisar

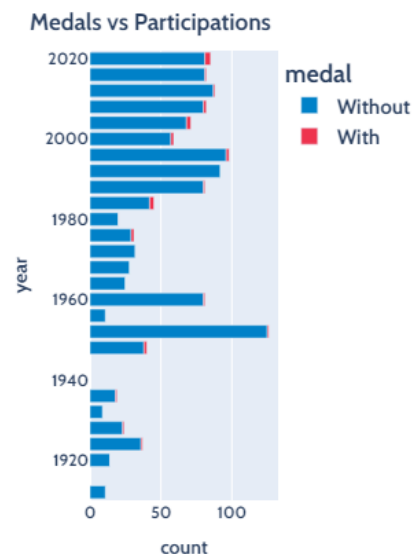


Figura 9. Comparação do número de medalhas arrecadadas pelo país em cada edição dos Jogos com o número de eventos participados

C. Vista das análises por modalidade

Se a opção selecionada for a de analisar os dados por modalidade, devem aparecer três visualizações. A principal deve ser uma raceplot (Figura 9) em que fica evidente a evolução do top 10 de medalhados numa determinada modalidade ao longo dos anos. Juntamente com esta visualização, do lado esquerdo, deve constar um gráfico de barras que expõe o top 10 de países mais medalhados nesse desporto (Figura 10) e outro que evidencia a evolução da participação de homens e mulheres nessa modalidade ao longo das edições (Figura 11). O layout desta vista deve ser o seguinte:

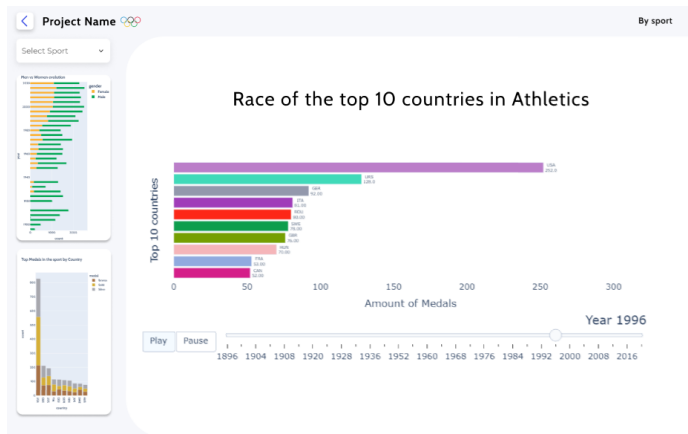


Figura 10. Vista das análises por modalidade

Top Medals by Country

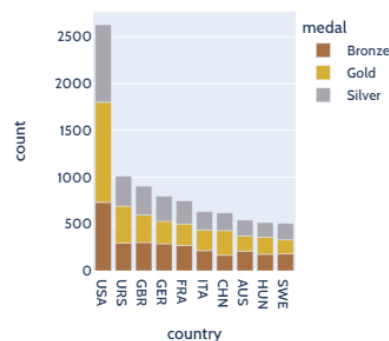


Figura 13. Top 10 de países na modalidade a analisar

As visualizações presentes são:

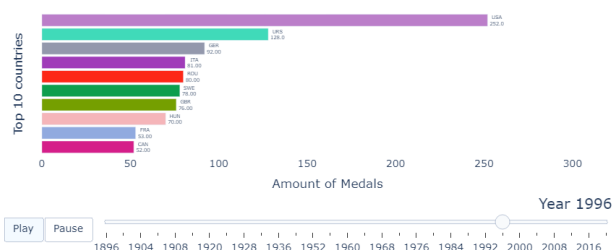


Figura 11. Raceplot dos países mais vencedores ao longo das edições na modalidade a analisar

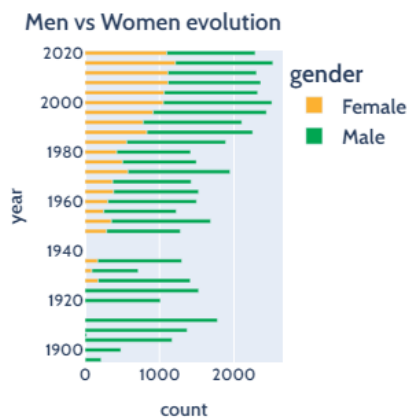


Figura 12. Comparação da participação de homens e mulheres nessa modalidade ao longo das edições das Olimpíadas

REFERÊNCIAS

- [1] “Jogos Olímpicos”, Wikipedia, 19-Jan-2023. Available: https://pt.wikipedia.org/wiki/Jogos_Ol%C3%ADmpicos. [Acedido: 24-Feb-2023]
- [2] Marcogdepinto, “Let’s discover more about the Olympic Games!”, Kaggle, 23-Sep-2018. Disponível: <https://www.kaggle.com/code/marcogdepinto/let-s-discover-more-about-the-olympic-games/notebook>. [Acedido: 24-Feb-2023]
- [3] J. Cheng, “Olympic historical dataset from Olympedia.org”, Kaggle, 26-Ago-2022. Disponível: https://www.kaggle.com/datasets/josephcheng123456/olympic-historical-dataset-from-olympediaorg?select=Olympic_Athlete_Bio.csv. [Acedido: 24-Feb-2023]
- [4] Raceplotly, PyPI. Disponível: <https://pypi.org/project/raceplotly/>. [Acedido: 24-Mar-2023]
- [5] Plotly, Plotly: Low-Code Data App Development. Disponível: <https://plotly.com/python/>. [Acedido: 24-Mar-2023]
- [6] plotly.graph_objects.Figure — 5.13.0 documentation, Plotly: Low-Code Data App Development. Disponível: https://plotly.com/python-api-reference/generated/plotly.graph_objects.Figure.html. [Acedido: 24-Mar-2023]
- [7] Zhu, A. (2022, 29 de março), How to Create Interactive Maps Using Python GeoPy and Plotly. Medium. Disponível: <https://towardsdatascience.com/work-with-geospatial-data-and-create-interactive-maps-using-geopy-and-plotly-28178>. [Acedido: 24-Mar-2023]
- [8] plotly.express.choropleth — 5.13.0 documentation, Plotly Help Center. Disponível: <https://plotly.github.io/plotly.py-docs/generated/plotly.express.choropleth.html>. [Acedido: 24-Mar-2023]
- [9] The D3 Graph Gallery – Simple charts made in d3.js, The D3 Graph Gallery – Simple charts made with d3.js. Disponível: <https://d3-graph-gallery.com/index.html>. [Acedido: 24-Mar-2023]
- [10] Data Viz Project, Data Viz Project. Disponível: <https://datavizproject.com/>. [Acedido: 24-Mar-2023]
- [11] <https://pkgstore.datahub.io/core/geo-countries/countries/archive/23f420f929e0e09c39d916b8aaa166fb/countries.geojson>. [Acedido: 24-Mar-2023]
- [12] [https://www.figma.com/proto/gUuP7vO7Ov52babz7Zkrzs/Analytics-Dashboard-\(Community\)?node-id=123-126&scaling=contain&page-id=0%3A1&starting-point-node-id=123%3A126](https://www.figma.com/proto/gUuP7vO7Ov52babz7Zkrzs/Analytics-Dashboard-(Community)?node-id=123-126&scaling=contain&page-id=0%3A1&starting-point-node-id=123%3A126). [Acedido: 24-Mar-2023]