

Olympics Dataset

Equipe: Liga Olímpica

Integrantes:

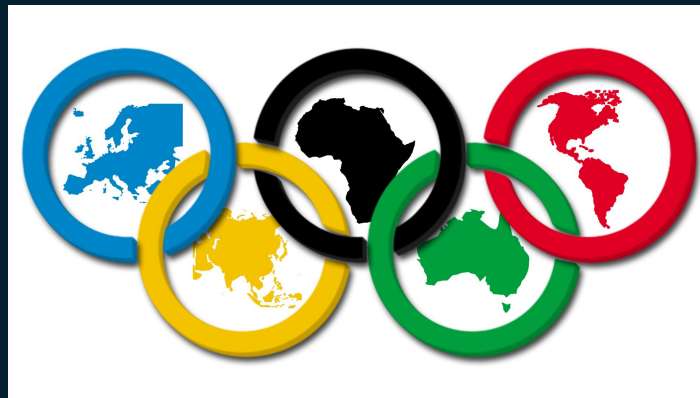
RA 213374 - Áureo Henrique e Silva Marques

RA 176566 - José Alexandre dos Santos Barros

RA 220407 - Lindon Jonathan Sanley dos Santos Pereira Monroe

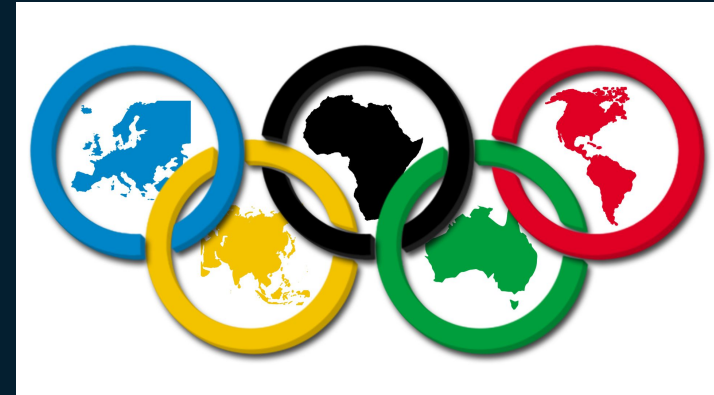
Tema

- Histórico dos Jogos Olímpicos nos últimos anos.
- Os Jogos Olímpicos, ou Olimpíadas, são o maior evento esportivo do mundo e, de 4 em 4 anos, reúnem milhares de atletas de vários países. Embora suas origens sejam da Grécia Antiga, as primeiras Olimpíadas ocorreram oficialmente em 1896, organizadas pelo Comitê Olímpico Internacional (COI) e, portanto, desse ano até hoje, tratam-se de mais de 30 edições dos jogos olímpicos.

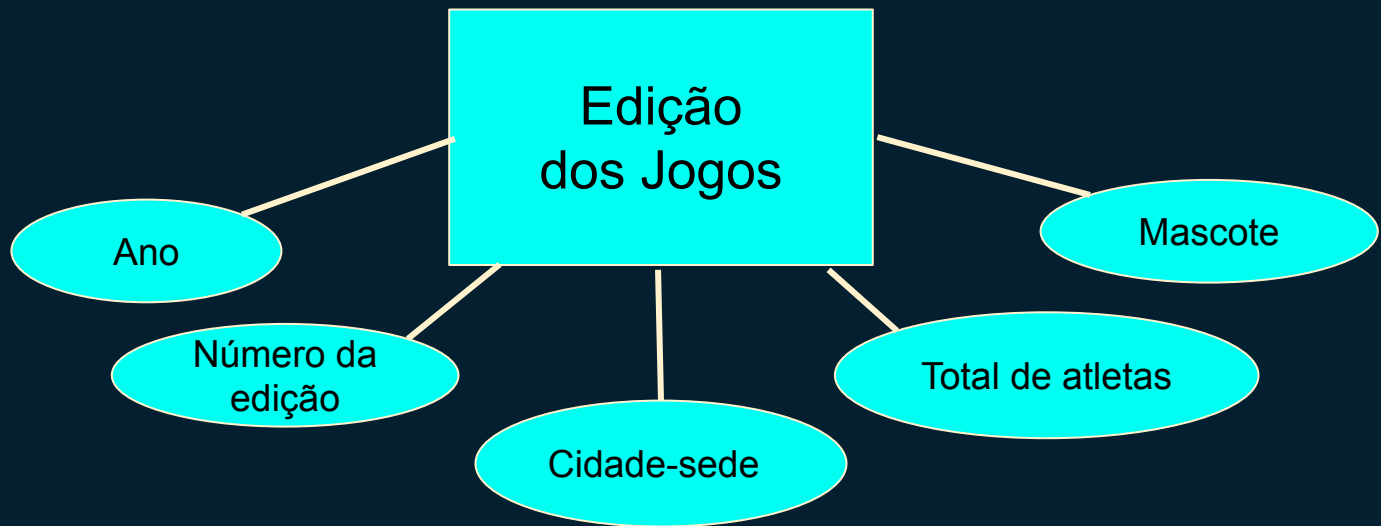


Tema

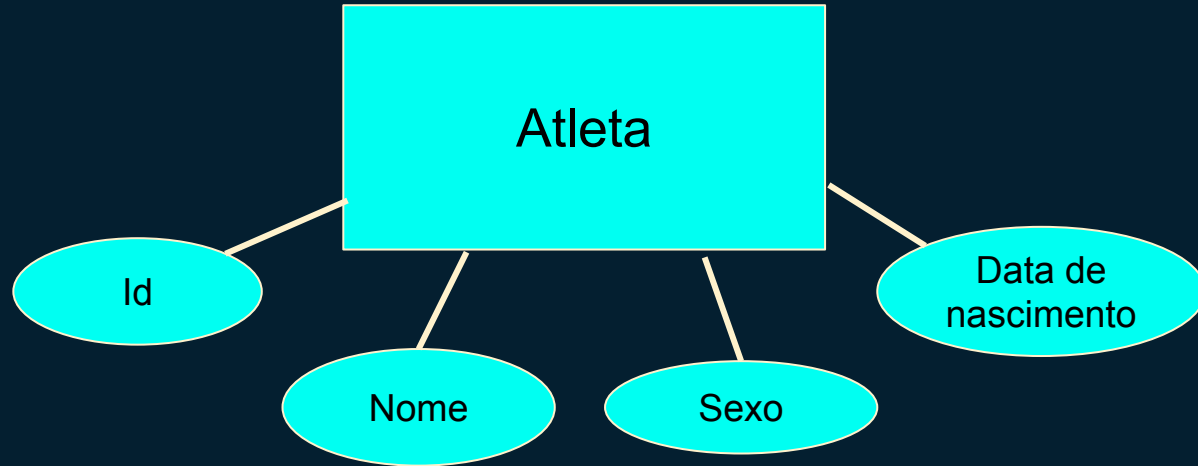
- Grande quantidade de informações sobre: atletas, países participantes, cidades-sede, medalhas, esportes, vencedores de cada modalidade, entre outros.
- É difícil encontrar bancos de dados atuais que possuam essas informações de forma centralizada e bem organizada.
- **Objetivo:** construir um dataset sobre os Jogos Olímpicos dos últimos anos que seja organizado e bem estruturado, permitindo diversos tipos de análises sobre o tema.



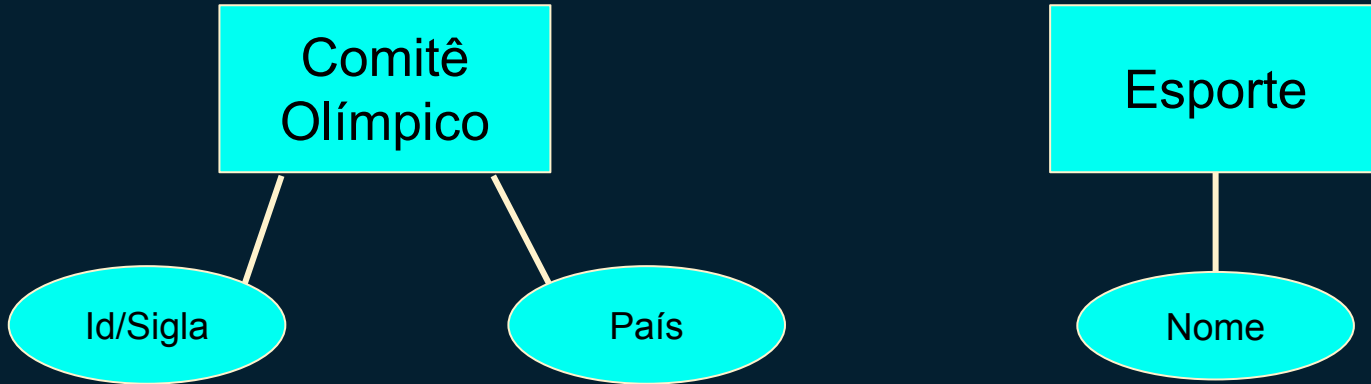
Modelo conceitual (entidade Edição dos Jogos)



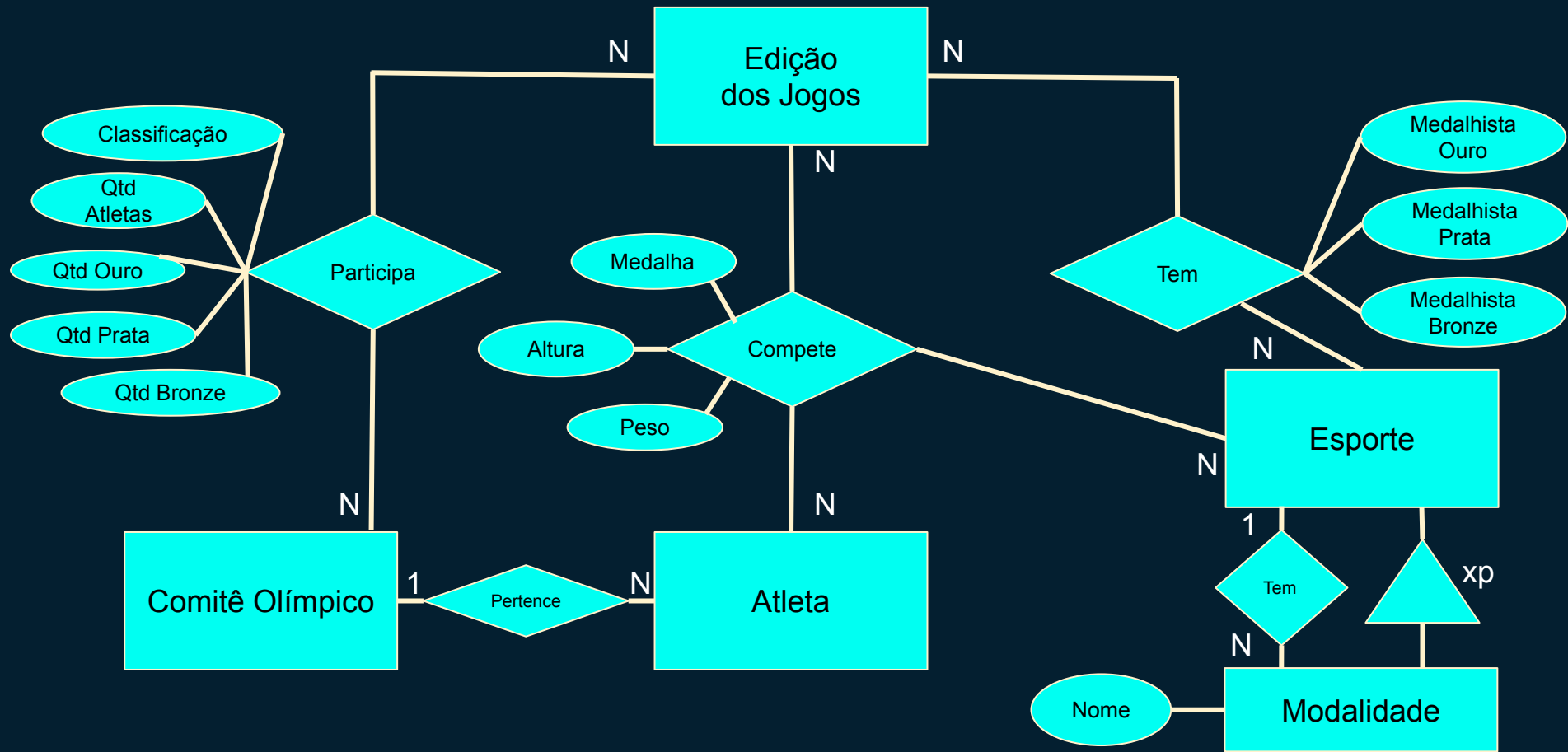
Modelo conceitual (entidade Atleta)



Modelo conceitual (entidade Comitê Olímpico e entidade Esporte)



Modelo conceitual (geral)



Fontes de dados

As fontes de dados que já pretendemos utilizar no dataset são as seguintes:

- Dataset histórico, uma tabela com dados das olimpíadas de 1896 a 2016. Foi criado a partir de dados do site www.sports-reference.com.

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results/discussion/69221>

- Dataset que consiste em uma tabela com dados específicos das olimpíadas no Rio de Janeiro em 2016.

<https://www.kaggle.com/rio2016/olympic-games>

- Dataset que consiste em uma tabela com dados específicos das olimpíadas de Tóquio em 2021.

<https://www.kaggle.com/arjunprasadsarkhel/2021-olympics-in-tokyo>

▲ Name	▲ Sex	▲ Age	▲ Height	▲ Weight	▲ Team	▲ NOC	▲ Games							
134732 unique values	M	73%	23	8%	NA	22%	NA	23%	United States	7%	USA	7%	2000 Summer	5%
	F	27%	24	8%	180	5%	70	4%	France	4%	FRA	5%	1996 Summer	5%
			Other (227521)	84%	Other (198453)	73%	Other (198616)	73%	Other (241281)	89%	Other (239505)	88%	Other (243515)	90%
A Dijiang	M	24	188	88	China	CHN	1992 Summer							
A Lamusi	M	23	178	68	China	CHN	2012 Summer							
Gunnar Nielsen Asby	M	24	NA	NA	Denmark	DEN	1928 Summer							

Modelo lógico e operações aplicadas aos dados

- **Modelos lógicos:** O modelo lógico relacional servirá para tratamento e melhor organização dos dados obtidos nas tabelas das fontes, além de utilizarmos o modelo de documentos. O modelo de redes ainda terá sua viabilidade em relação ao projeto analisada.
- **Operação de integração:** Os dados contidos nas múltiplas fontes, como as diferentes tabelas dos datasets, serão integrados de forma a criar um dataset mais rico.
- **Operação de transformação:** Os dados obtidos serão transformados através da sua reorganização, de modo a ficar de acordo com o planejado no modelo conceitual.

Conjunto inicial de perguntas

- Quais os países que mais ganharam medalhas e os países que menos ganharam medalhas em uma determinada Olimpíada?
- Qual o número médio de medalhas de um país nas Olimpíadas que ele participou?
- Qual o número de atletas por país em uma determinada olimpíada?
- Quais foram os países ganhadores de medalha de ouro no esporte X nas últimas 5 olimpíadas?
- Qual a proporção de atletas do sexo masculino e do sexo feminino participando nos Jogos Olímpicos?
- Em quantas Olimpíadas um determinado atleta participou e quantas medalhas ele ganhou?
- Quais países que mais trazem atletas para os Jogos Olímpicos?
- Para um determinado esporte, existe algum país que constantemente está no pódio?
- Para os países que trazem um número reduzido de atletas, em quais esportes eles costumam participar?

The image features a dark navy blue background. In the top-left and bottom-left corners, there are overlapping, semi-transparent geometric shapes in shades of green, blue, orange, and pink. Similarly, in the top-right and bottom-right corners, there are overlapping, semi-transparent geometric shapes in shades of green, blue, orange, and pink. The word "Obrigado!" is centered in the middle of the image in a white, bold, sans-serif font.

Obrigado!