



Folha de Exercícios n.º PL02	
Unidade Curricular:	<b>Análise de Dados</b>
Ano Letivo:	2019/2020 - 1.º Semestre

## 1 Modelo Dimensional de Eventos dos Jogos Olímpicos

O ficheiro disponibilizado juntamente com esta ficha prática laboratorial, nomeadamente `athlete_events.csv`, contém dados reais históricos sobre os Jogos Olímpicos modernos, incluindo todos os Jogos Olímpicos desde os Jogos Olímpicos de Atenas de 1896 até aos Jogos Olímpicos de Rio de 2016. Os dados de 1000 registos foram extraídos para um ficheiro no formato `.csv` a partir de um *dataset* inicial com 271116 linhas.

No ficheiro `athlete_events.csv`, cada linha corresponde a um atleta competindo num determinado evento olímpico. A informação representada inclui 14 colunas, nomeadamente: `id` (identificador único de cada evento), `id_athlete` (identificador único do atleta), `name_athlete`, `sex` (*M* ou *F*), `age`, `height` (em centímetros), `weight` (em quilogramas), `team` (nome da equipa), `games` (ano e temporada), `ano`, `temporada` (*Summer* ou *Winter*), `cidade` (cidade anfitriã), `evento` (especificação do desporto) e `medalha` (*Gold*, *Silver*, *Bronze* ou *NA*).

Note-se que os Jogos de Inverno e os Jogos de Verão foram realizados no mesmo ano até 1992. Depois de 1992, os Jogos Olímpicos foram escalonados de tal forma que os Jogos de Inverno ocorrem cada quatro anos começando em 1994, depois os Jogos de Verão em 1996, depois os Jogos de Inverno em 1998, e assim por diante. Um erro comum que poderia cometer ao analisar este *dataset* seria assumir que os Jogos de Inverno e os Jogos de Verão sempre foram escalonados.

Assim, este conjunto de dados representa uma oportunidade única para fazer questões sobre como os Jogos Olímpicos evoluíram ao longo do tempo, incluindo sobre a participação e o desempenho de mulheres, das diferentes nações e dos diferentes desportos e eventos, entre outros.

Os profissionais de tecnologias de informação de uma empresa pretendem remodelar a organização da informação em questão num modelo dimensional baseado no esquema em estrela.

Com base no caso apresentado, pretende-se que:

1. Crie um novo *schema* no MySQL Workbench denominado “Ficha2”.
2. Faça o *import* dos dados no ficheiro `athlete_events.csv` para uma nova tabela no *schema* criado na alínea anterior. No processo, uma tabela denominada “`athlete_events`” deverá ser criada e povoada corretamente com os dados do ficheiro (Table Data Import Wizard).
3. Analise a estrutura da tabela `athlete_events` e, conseqüentemente, define um modelo dimensional no formato de esquema em estrela. O modelo deverá ter uma tabela de factos e as respetivas tabelas de dimensão ligadas à tabela de factos.
4. Construa o modelo dimensional definido na alínea anterior no MySQL Workbench (*EER diagram*).
5. Faça a conversão do modelo lógico criado para o respetivo modelo físico para o *schema* Ficha2 (Database > Forward Engineer).
6. Povoie todas as tabelas do modelo dimensional (tabela de factos e tabelas de dimensão) em SQL a partir da tabela `athlete_events`. É de relembrar que as tabelas de dimensão deverão ser povoadas antes das tabelas de factos.