



Folha de Exercícios n.º PL03	
Unidade Curricular:	Análise de Dados
Ano Letivo:	2019/2020 - 1.º Semestre

1 Esquema em Estrela vs. Esquema em Floco de Neve

O assassinato em 2014 de Michael Brown em Ferguson, Missouri, Estados Unidos da América (EUA), iniciou um movimento de protesto que culminou com o *Black Lives Matter* e um foco maior na responsabilidade dos polícias em todo o país.

Desde o dia 1 de Janeiro de 2015, *The Washington Post* tem vindo a recolher dados numa base de dados relativos a todos os disparos fatais nos EUA por um polícia durante o seu cumprimento de dever legal.

É interessante referir que é difícil encontrar dados confiáveis antes do dia 1 de Janeiro de 2015 uma vez que este tipo de acontecimentos não era documentado de forma abrangente, e estatísticas sobre a brutalidade policial estão ainda muito menos disponíveis. Como resultado, um grande número deste tipo de casos não está relatado.

The Washington Post está a recolher mais de uma dúzia de detalhes sobre cada assassinato, incluindo a raça, a idade e o género do falecido, se a pessoa estava armada e se a vítima estava num estado de crise (saúde mental).

O ficheiro disponibilizado juntamente com esta ficha prática laboratorial, nomeadamente `police_killings_us.csv`, contém os dados reais dessa recolha realizada pelo *The Washington Post*. Cada linha do ficheiro corresponde a um disparo fatal por um polícia nos EUA desde 2015. A informação representada inclui 14 colunas, nomeadamente: `id` (identificador único de cada disparo fatal), `name` (da vítima), `date` (da morte da vítima), `manner_of_death`, `armed`, `age`, `gender`, `race`, `city`, `state`, `signs_of_mental_illness`, `threat_level`, `flee` e `body_camera`.

Assim, este conjunto de dados representa uma oportunidade única para fazer questões relevantes sobre a brutalidade policial nos últimos anos nos EUA.

Com base no caso apresentado, pretende-se que:

1. Analise a estrutura da tabela `police_killings_us.csv` e, conseqüentemente, defina um modelo dimensional no formato de esquema em estrela.
2. Analise a estrutura da tabela `police_killings_us.csv` e, conseqüentemente, defina um modelo dimensional no formato de esquema em floco de neve.
3. Construa cada um dos modelos dimensionais definidos nas alíneas anteriores no MySQL Workbench (*EER diagram*).
4. Descreva as vantagens e as desvantagens entre os dois diferentes tipos de modelo dimensional definidos.
5. Defina dez questões de interesse que poderia colocar a esta base de dados. Justifique a relevância de cada uma das questões definidas.