

Análise Exploratória de Dados

A base de dados **BD_Netflix.csv** tem informação relativa a programas de televisão (TV) e filmes disponíveis na Netflix. Esta base de dados permite responder a diversas perguntas e retirar diversas conclusões, como por exemplo:

- Qual foi a evolução dos programas de TV disponíveis na Netflix na última década? E dos filmes?
- Quais são as principais categorias de programas de TV/filmes?
- Qual é a duração média dos programas de TV/filmes? Em que categorias?
- Quem são os principais realizadores dos programas de TV/filmes?
- Quais são os países com maior número de programas de TV/filmes?
- Que tipo de conteúdos estão disponíveis dos diferentes países?

Antes de começar a utilizar uma base de dados é necessário fazer uma primeira análise exploratória dos dados. Neste trabalho pretende-se que comece por conhecer a base de dados, realizando quadros e gráficos simples, utilizando Python. No final, deverá ser capaz de responder às perguntas indicadas acima e propor outras perguntas interessantes para a exploração da base de dados. Utilize Python para completar os seguintes exercícios.

Exercício 1

Analise a qualidade e informação disponível na base de dados. Verifique que variáveis estão disponíveis, valores em falta, quais são as suas estatísticas básicas (valores médios, máximos, variância, etc...).

Verifique qual é a proporção entre programas de TV/filmes. Faça um “pie chart” ou um gráfico de barras.

Exercício 2

Verifique a evolução temporal de novos programas de TV/filmes nos últimos anos. Faça um gráfico de frequência ao longo do tempo, discriminando “programas de TV” e “filmes”, por ano de lançamento (release_year).

Exercício 3

Identifique quem são os principais realizadores, nas diversas categoriais. Faça um gráfico de frequências com os 8 principais realizadores, para programas de TV e filmes.

Exercício 4

Implemente um programa que permita responder às restantes questões acima.

Exercício 5

Sugira mais 2 ou 3 perguntas, e apresente a respetiva solução.