

Introdução à Ciência dos Dados – Prova Prática 01

Professor: Fabrício A. Silva

Valor: 20 pontos

Data: 29/10/2020 (Entrega via PVANet)

Observações: é permitido usar os arquivos e códigos já desenvolvidos ao longo da disciplina, e consultar qualquer conteúdo na Internet. Não é permitida nenhuma interação com colegas.

Leia o dataset “*Filmes-Prova01.csv*” (PVANet). Entregar um zip com os arquivos em format .ipynb e html. Esse conjunto de dados possui a lista de filmes de quatro plataformas de *Streaming* (Netflix, Amazon Prime, Disney+ e Hulu), com algumas de suas características e suas notas no IMDb. Com base nesses dados, resolva:

1. (0,5 ponto) Quantos filmes cada uma das 4 plataformas possui? Faça um gráfico de barras para ilustrar esses valores.
2. (0,5 ponto) Quantos filmes estão em mais de uma plataforma?
3. (1 ponto) Qual a plataforma que possui a maior média de nota IMDb?
4. (1 ponto) Qual o percentual de filmes de cada plataforma com nota IMDb maior que 8?
5. (1 ponto) Se uma pessoa é uma apreciadora de filmes clássicos antigos, qual plataforma você mais recomenda? Justifique.
6. (1 ponto) Quantas categorias de classificação etária existem? Faça um gráfico de barras com a quantidade de filmes por classificação etária.
7. (1 pontos) Qual plataforma você mais indicaria para uma criança? Justifique.
8. (2 pontos) Faça um gráfico de linhas com a quantidade de filmes disponíveis por ano de lançamento. Qual o ano que possui mais filmes, e quantos filmes foram lançados nesse ano? Qual a média de lançamentos de filmes por ano? Trace uma linha reta no gráfico com a média, e responda em quais anos foram lançados mais filmes do que a média.
9. (3 pontos) Faça o *Boxplot* da duração dos filmes. Em seguida, remova os *outliers* que achar necessário, e faça o histograma da duração. Ajuste o parâmetro "bins" se necessário, para visualizar melhor. Em seguida, responda se os valores de duração seguem a distribuição Normal. Justifique.
10. (3 pontos) Quantos gêneros distintos existem? Quantos filmes de cada gênero existem?
11. (3 pontos) Qual o diretor possui a melhor media de nota IMDb dos seus filmes?
12. (3 pontos) Utilize regras de associação para responder às seguintes questões:
 - a) Qual o conjunto de itens de gêneros com mais de um gênero que aparece em mais filmes?
 - b) Qual o percentual de filmes em que esse conjunto de itens aparece?
 - c) Considerando uma pessoa que gosta de filmes de "Crime", quais outros gêneros você recomendaria a essa pessoa? Justifique sua resposta.