



INSTITUTO FEDERAL DE
EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
São Paulo
Campus Campinas

Instituto Federal de Educação, Ciência e Tecnologia

Câmpus Campinas

D1AED – Análise Estatística para Ciência de Dados

Professores: Ricardo Sovat (sovat@ifsp.edu.br)
Samuel Martins (samuel.martins@ifsp.edu.br)

Atividade em Dupla 1

1. Especificação

Nesta atividade, cada dupla deverá realizar uma **análise exploratória** para o conjunto de dados *Movie Lens (small)*: <https://grouplens.org/datasets/movielens>, a partir de um conjunto de perguntas e hipóteses pré-estabelecidas e propostas pela dupla.

Seguem os critérios a serem avaliados. Cada critério tem um conjunto de pontos que servirão como um guia para seu desenvolvimento.

- Descrição da base de dados escolhida: **[1 ponto]**
 - O que significa cada linha de cada arquivo?
 - Quais são os principais atributos (colunas) e seus tipos?
- Preparação da base de dados: **[1 ponto]**
 - Teve dificuldades para preparar o dataset para uso (p. ex., nomes ruins para as colunas, arquivos dos datasets não possuíam boa organização)?
 - Aplicou algum pré-processamento nos dados?
 - ◆ Outliers/ruídos, dados faltantes, conversão de tipos de dados, etc.
- Análise exploratória: **[6 pontos]**
 - Respostas e discussão dos achados das seguintes perguntas/hipóteses propostas:
 - ◆ Pergunta 1 – Existe alguma correlação entre o ano de lançamento do filme e sua avaliação média (média de ratings)?
 - ◆ Pergunta 2 – Qual o grupo mais disperso, o de usuários que avaliaram ou o de usuários que rotularam filmes?
 - ◆ Pergunta 3 – Qual a categoria rotulada mais frequente?
 - ◆ Pergunta 4 – Como estão distribuídas as avaliações ao longo do tempo?
 - ◆ Pergunta 5 – pergunta sugerida pela dupla
 - ◆ Pergunta 6 – pergunta sugerida pela dupla
 - ◆ Pergunta 7 – pergunta sugerida pela dupla
 - ◆ Pergunta 8 – pergunta sugerida pela dupla
 - ◆ Pergunta 9 – pergunta sugerida pela dupla
 - ◆ Pergunta 10 – pergunta sugerida pela dupla
- Relatório (Notebook): **[1 ponto]**
 - Organização do relatório;
 - Clareza na apresentação dos textos e códigos;
 - Discussão sucinta dos achados

- Vídeo de apresentação da análise exploratória (**até 10 min**): [1 ponto]
 - Ambos os integrantes devem participar;

2. Submissão (03/06/21)

A submissão da análise exploratória será feita em tarefa específica no Moodle da disciplina. **Um membro** da dupla poderá enviar um jupyter notebook (.ipynb) ou o link do repositório online com o código (ex., Google Colab, GitHub, Kaggle). No caso dos links para repositórios ou plataformas online, serão considerados apenas aqueles com atualização até o prazo de entrega desta atividade.

3. Material de Apoio

Dataset a ser usado:

- Movie Lens (latest small): <https://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

Visualizações:

- <https://datavizcatalogue.com/index.html>
- <https://www.python-graph-gallery.com/>