



T2 (peso 3)

As notas da disciplina de Ciência dos dados 2025.2 serão atribuídas a partir de trabalhos desenvolvidos em grupo. Este segundo trabalho tem como objetivo desenvolver as habilidades de **Análise Exploratória de Dados** (*Exploratory Data Analysis*, EDA) e **Visualização de Dados**. Sigam as seguintes instruções abaixo:

- A entrega deve ser feita pelo mesmo grupo estabelecido no trabalho 1.
- O trabalho tem dois principais objetivos:
 1. Apresentar e caracterizar o conjunto de dados escolhido através de uma análise exploratória dos dados.
 2. Formular e responder perguntas sobre os dados através de análises e visualizações de dados.
- Na parte da **Análise Exploratória dos Dados**, é esperado:
 - Apresentar os dados escolhidos, de onde eles foram extraídos, quantas instâncias e atributos ele tem, e em qual período de tempo, identificar se há alguma licença de uso associado aos dados, se há questões éticas e/ou privacidade que possam ser levantadas. Justifique.
 - Descrever todos os processamentos feitos nos dados (dados faltantes, transformações de tipo, composições, correções, normalizações, *outliers*, verificações de consistência interna).
 - Sintetizar através de estatísticas e visualizações simples os atributos do conjunto de dados. Em caso de MUITOS atributos, selecione o subconjunto que julgar mais interessante.
 - Buscar apresentar correlações entre atributos, estatísticas básicas, distribuições, relações temporais e quaisquer dúvidas acerca dos dados nesta etapa.
 - Entender, com certa profundidade, através de diferentes técnicas os dados que estão disponíveis, sendo esse o principal objetivo desta etapa.
- Na parte de **Visualização dos dados** é esperado:
 - Crie entre 6-8 perguntas e hipóteses acerca dos dados, buscando sempre relacionar algum **propósito a suas perguntas**. Por exemplo, *ao responder essa pergunta, quais benefícios sociais ou financeiros podem ser alcançados com esse insight? Quais ações e decisões podem ser tomadas com essas informações adquiridas?*
 - Tente responder cada uma delas através de visualizações que **valorizem** o que foi percebido para responder a pergunta.

- Busque aplicar diferentes visualizações para as perguntas feitas, mas sempre buscando aplicar visualizações consistentes com os dados.
- Todas as visualizações devem possuir título (subtítulo quando pertinente) com informações que respondam perguntas como “o quê?”, “onde?”, “quando?”.
- Atenção aos rótulos e unidades: eixos claros, escalas apropriadas, legendas breves, notas quando houver filtros.
- Busque manter consistência visual: por exemplo, se uma categoria aparece em diversas visualizações (país, sexo, produto, etc.) procure sempre apresentá-los com os mesmos marcadores e (principalmente) canais nas diferentes visualizações.
- Será necessário apresentar pelo menos um Notebook EDA (1 principal, claro e linear) com conclusões e comentários marcadas ao longo do arquivo, assim como os códigos para criação das visualizações.
- Relatório curto (2–4 págs) com um resumo geral do que foi feito, com objetivo, dados utilizados, e principais achados (3–5 gráficos), assim como apontar limitações e recomendações. Adicione também quais foram os principais desafios e aprendizados durante o desenvolvimento do trabalho.
- Cada grupo deve apresentar os artefatos do trabalho em uma apresentação de até 10 minutos. A apresentação é livre (você pode fazer slides ou apresentar diretamente do código e/ou notebooks), mas é necessário mostrar os notebooks e códigos utilizados no desenvolvimento do trabalho, assim como o e relatório gerado, durante a apresentação. **Tudo deve estar disponível no Github do projeto.**
- O trabalho valerá 10 pontos divididos nos seguintes critérios:
 1. Profundidade da EDA (20%) - quanto claro ficou o entendimento sobre os dados apresentados.
 2. Qualidade das visualizações & boas práticas (20%) - uso de visualizações apropriadas, escolha de marcadores e canais, etc.
 3. Storytelling, insights e recomendações (10%) - Como a apresentação ocorreu e como os achados foram apresentados.
 4. Reprodutibilidade & organização do projeto (10%) - Se as bibliotecas estão no `requirements.txt`; códigos consolidados estão no `src`; se as figuras e relatórios estão nas pastas apropriadas.
 5. Relatório (10%).
 6. Resposta às perguntas feitas após a apresentação (20%). As perguntas podem ser relacionadas ao trabalho em si ou sobre conceitos apresentados em sala. Perguntas poderão ser feitas diretamente a integrantes específicos.