Bacharelado em Engenharia de Software

Disciplina: Ciência de Dados

Você recebeu o arquivo `atividade3_dataset.csv` de um cliente, mas ele contém sérios problemas de qualidade. Sua tarefa é **preparar essa base para análises futuras**.

1. **Leitura e inspeção**
    - Base de dados disponível em: atividade3_dataset;
    - Carregue a base e verifique: tipos de dados, quantidade de nulos e duplicados.
2. **Tratamento de valores ausentes**
    - Defina e justifique diferentes estratégias:
        - Remover linhas/colunas?
        - Preencher com média/mediana/moda?
        - Usar um valor padrão (ex: “Desconhecido”)?
    - Explique sua escolha para cada coluna.
3. **Detecção e remoção de duplicados**
    - Quantas linhas duplicadas existem?
    - Mantenha apenas as válidas.
4. **Correções adicionais**
    - Padronize valores inconsistentes (ex.: "SP" vs. "São Paulo").
    - Converta colunas que estão com tipo errado (números armazenados como texto, datas como strings).
5. **Exploração inicial após limpeza**
    - Estatísticas descritivas (média, mediana, desvio padrão).

- Pelo menos **2 gráficos** (distribuição e correlação).

6. **Entrega**

    - Enviar um notebook (`.ipynb`) contendo o código **e explicações em texto** sobre cada decisão.

    - Exportar também a base limpa como `base_limpa.csv`.
    - Ambos os arquivos serão recebidos até dia **04/09**, exclusivamente pelo seguinte link: Formulário de entrega da atividade

Desafio:

- Criar pelo menos **uma nova coluna derivada** a partir dos dados originais.

- Exemplo: faixa etária, classificação de renda, indicador binário (acima/abaixo da média).