

Atividade: Criação, Correção e Visualização de Datasets com Pandas e Matplotlib

Objetivo

Desenvolver a capacidade de:

- Criar datasets de forma intencional;
 - Identificar e corrigir problemas comuns em dados usando **pandas**;
 - Enriquecer datasets com novas colunas derivadas;
 - Explorar e comunicar informações por meio de **gráficos com matplotlib**.
-

Parte 1 – Organização da Atividade

- A turma deve ser organizada em **duplas**.
 - Apesar de estarem em dupla, **cada aluno inicia o trabalho individualmente**, criando o seu próprio dataset.
 - Posteriormente, os datasets serão **trocados entre os colegas da dupla**.
-

Parte 2 – Criação Individual do Dataset

Cada aluno deverá **gerar um dataset fictício**, utilizando uma IA ou criando manualmente, seguindo as regras abaixo.

1 Tema do Dataset

Escolha **um contexto simples e realista**, como por exemplo:

- Vendas de uma loja

- Registros escolares
 - Chamados de suporte
 - Pedidos de delivery
 - Dados de uma academia, hospital, transporte, etc.
-

2 Estrutura do Prompt para Criação do Dataset

O dataset deve ser criado a partir de um prompt seguindo **obrigatoriamente esta estrutura**:

Prompt base (modelo):

“Crie um dataset fictício no formato de tabela (CSV ou tabela estruturada) sobre **[tema escolhido]**, contendo pelo menos **X linhas e Y colunas**.”

O dataset deve conter **intencionalmente pelo menos 2 problemas de dados**, como por exemplo:

- valores ausentes (NaN),
- tipos de dados incorretos,
- datas em formatos diferentes,
- valores duplicados,
- erros de digitação em categorias.

Além disso, o dataset deve permitir a criação de **pelo menos 2 colunas derivadas (enriquecimento)**, como:

- totais,
- médias,
- classificações,
- extração de partes de datas,

- conversões ou faixas de valores.

Não explique os problemas inseridos no dataset.”



Importante:

O aluno **não deve indicar ao colega quais são os problemas** inseridos no dataset.

3 Requisitos do Dataset

O dataset criado deve:

- Ter **dados suficientes para análise** (mínimo recomendado: 20 linhas);
- Possuir **pelo menos 2 problemas de dados** que precisem ser corrigidos com pandas;
- Permitir **pelo menos 2 enriquecimentos de colunas**;
- Estar salvo em formato utilizável (CSV ou DataFrame).

Parte 3 – Troca de Datasets

- Após a criação, os alunos devem **trocar seus datasets com o colega da dupla**.
- A partir desse momento, cada aluno trabalhará **apenas com o dataset recebido**.

Parte 4 – Correção e Leitura do Dataset (Pandas)

Com o dataset recebido, o aluno deverá:

1. **Carregar o dataset com pandas**.
2. Identificar e corrigir **pelo menos 2 problemas**, como:

- Ajuste de tipos de dados (`astype`, `to_datetime`, etc.);
 - Tratamento de valores nulos;
 - Remoção ou correção de duplicatas;
 - Padronização de textos ou categorias.
3. Criar **pelo menos 2 novas colunas derivadas**, utilizando operações como:
- Cálculos matemáticos;
 - Condições (`if`, `np.where`, `apply`);
 - Manipulação de datas;
 - Classificações ou agrupamentos.
4. Realizar uma **leitura exploratória dos dados**, utilizando:
- `head()`, `info()`, `describe()`,
 - contagens, médias ou agrupamentos simples.
-

Parte 5 – Visualização de Dados com Matplotlib

Com o dataset corrigido e enriquecido, o aluno deverá criar **no mínimo 4 gráficos diferentes**, utilizando **matplotlib**, respeitando os critérios abaixo:

Requisitos dos Gráficos

- Os gráficos devem ser **inspirados nos dados do dataset**;
- Utilizar **tipos variados**, como por exemplo:
 - gráfico de linha,
 - gráfico de barras,

- gráfico de dispersão,
 - gráfico de barras agrupadas.
 - Cada gráfico deve conter:
 - título,
 - rótulos dos eixos,
 - legenda (quando aplicável).
 - Sempre que possível, aplicar ajustes visuais:
 - cores,
 - marcadores,
 - tamanhos,
 - rotação de rótulos.
-

Entrega

A atividade deverá ser entregue até o dia **19/02/2026 (Quinta-feira)**.

As duplas deverão enviar um **arquivo compactado** (ZIP ou similar) por e-mail, contendo os seguintes itens de ambos os colegas:

1. Os dois **datasets originais** (recebidos na troca).
2. Os dois **códigos Python** (ou notebooks) contendo a leitura, correção, enriquecimento e geração dos gráficos.
3. As **versões corrigidas/enriquecidas dos datasets** (se exportadas).

Endereço de Envio: tarikponciano@ce.senac.br

Assunto do E-mail: Atividade Pandas e Matplotlib

Corpo do E-mail (Obrigatório): Mencionar o **nome completo dos dois integrantes** da dupla.

Entregáveis

Cada aluno deverá entregar:

1. O **dataset recebido** (original).
2. O **código em Python** contendo:
 - leitura do dataset,
 - correções realizadas,
 - enriquecimento das colunas,
 - geração dos 4 gráficos.
3. (Opcional) Um breve comentário explicando:
 - quais problemas foram encontrados,
 - quais colunas foram enriquecidas,
 - o que cada gráfico representa.