

Guia: Gerador de Dataset Simulado de Clientes e Compras

Guia: Gerador de Dataset Simulado de Clientes e Compras

O que o código faz?

Este código gera um dataset simulado contendo informações detalhadas sobre clientes, suas características demográficas e o histórico de compras, incluindo produtos adquiridos, avaliações, métodos de pagamento e datas. O dataset é salvo como um arquivo CSV que pode ser utilizado para análises de comportamento de clientes, estudos de marketing, treinamento de modelos de machine learning ou outras finalidades.

Estrutura do Dataset Gerado

O dataset gerado possui as seguintes colunas:

1. **Nome**: Nome fictício do cliente, atribuído com base no gênero (ex.: "Luiz" para masculino, "Carla" para feminino).
2. **ID**: Um identificador único para cada cliente.
3. **Idade**: A idade do cliente, gerada aleatoriamente entre 18 e 60 anos.
4. **Gênero**: O gênero do cliente, podendo ser "Masculino" ou "Feminino".
5. **Cidade**: A cidade onde o cliente reside (escolhida entre Fortaleza, São Paulo, Rio de Janeiro, Salvador e Belo Horizonte).
6. **Estado**: O estado correspondente à cidade (ex.: Fortaleza pertence ao estado "Ceará").
7. **Produto**: Um produto comprado pelo cliente (escolhido aleatoriamente a partir de uma lista de produtos).
8. **Valor**: O preço do produto em reais, gerado aleatoriamente entre R\$5,00 e R\$200,00.
9. **Data_Compra**: A data da compra, simulada entre 01/01/2023 e 31/12/2023.
10. **Avaliacao**: Uma avaliação do produto, podendo ser "Excelente", "Boa", "Neutra", "Ruim" ou

Guia: Gerador de Dataset Simulado de Clientes e Compras

"Péssimo".

11. ****Pagamento****: O método de pagamento usado na compra, como Pix, Crédito, Débito ou Dinheiro.

Cada cliente pode ter várias compras associadas, gerando múltiplas linhas no dataset para um mesmo cliente.

Passo a Passo do Código

1. ****Configuração Inicial****:

- O número de clientes é definido (`n_linhas`).
- IDs únicos são gerados para cada cliente.
- Listas com opções de idades, gêneros, cidades e produtos são preparadas.

2. ****Geração de Dados Demográficos****:

- ****Idade****: Atribuída aleatoriamente entre 18 e 60 anos.
- ****Gênero****: Determinado aleatoriamente como "Masculino" ou "Feminino".
- ****Cidade****: Selecionada aleatoriamente de uma lista pré-definida, com o estado correspondente atribuído automaticamente.

3. ****Criação de Nomes****:

- Nomes fictícios são atribuídos com base no gênero:
- Masculino: Ex.: "Luiz", "João", "Carlos".
- Feminino: Ex.: "Carla", "Maria", "Ana".

4. ****Simulação de Compras****:

Guia: Gerador de Dataset Simulado de Clientes e Compras

- Para cada cliente, é gerado um número aleatório de produtos adquiridos (até 6).
- Produtos são escolhidos aleatoriamente de uma lista.
- Preços dos produtos são gerados de forma aleatória, variando entre R\$5,00 e R\$200,00.
- Datas de compras são simuladas ao longo de 2023.
- Avaliações e métodos de pagamento são atribuídos de forma aleatória.

5. **Criação do Dataset**:

- Cada compra é registrada como uma linha no dataset, incluindo todas as informações demográficas do cliente e os detalhes da compra.

6. **Salvar em CSV**:

- O dataset gerado é salvo em um arquivo CSV no caminho especificado (`file_path`).
- O CSV pode ser aberto em softwares como Excel, Google Sheets ou lido diretamente em outras aplicações Python.

Como Usar Este Código

1. **Pré-requisitos**:

- Instalar a biblioteca `pandas` para manipulação de dados.
- Certifique-se de ter permissões para salvar arquivos no caminho especificado.

2. **Configurações Personalizáveis**:

- `n_linhas`: Altere o número de clientes para gerar datasets maiores ou menores.
- `produtos_variados`: Adicione ou remova produtos da lista para adequar o dataset ao seu cenário.
- `avaliacoes` e `pagamentos`: Personalize as opções de avaliações e métodos de pagamento.

Guia: Gerador de Dataset Simulado de Clientes e Compras

3. ****Executar o Código****:

- Após configurar as opções desejadas, execute o código para gerar o arquivo CSV.

4. ****Usar o Dataset****:

- O arquivo gerado estará pronto para ser utilizado em qualquer ferramenta que suporte arquivos CSV.

Exemplo de Linha Gerada no CSV

```
```csv
```

```
Nome,ID,Idade,Gênero,Cidade,Estado,Produto,Valor,Data_Compra,Avaliacao,Pagamento
```

```
Luiz,101,35,Masculino,Fortaleza,Ceará,Tênis,R$120.50,2023-07-15,Boa,Pix
```

```
Carla,102,28,Feminino,Salvador,Bahia,Camiseta,R$75.99,2023-03-21,Excelente,Crédito
```

```
```
```

Possíveis Aplicações

1. Treinamento de modelos de aprendizado de máquina.
2. Simulação de cenários de comportamento do consumidor.
3. Análise exploratória e prática com dados para aprendizado em Python.
4. Estudo de tendências de compras e segmentação de clientes.

```
---
```