

## Organização de Dados para Análise

A empresa fictícia "SmartSolutions" está desenvolvendo um novo sistema de gerenciamento de clientes e vendas. No entanto, os dados fornecidos pelos setores estão desorganizados e inconsistentes.

Sua equipe foi contratada para organizar os dados para análise e modelagem.

Vocês receberão uma planilha com dados brutos contendo erros, duplicações, formatações inconsistentes e dados ausentes.

1. Aplicar um checklist de organização de dados
2. Identificar e corrigir inconsistências, como:
  - Dados faltantes
  - Duplicatas
  - Formatação incorreta (datas, números, textos)
3. Padronizar os dados de forma coerente
4. Estruturar os dados em tabelas lógicas separadas:
  - Clientes
  - Vendas
5. Criar um mini dicionário de dados, explicando os campos
6. Entregar:
  - Planilha corrigida
  - Checklist preenchido
  - Dicionário de dados

## Modelo de checklist

### Informações do Projeto

- Nome: \_\_\_\_\_
- Projeto/Situação: \_\_\_\_\_
- Data: //\_\_\_\_\_

### Etapas do Checklist

Etapa	Descrição da Tarefa	Status (✓/✗/□)	Observações
1	Identificar fontes de dados	<input type="checkbox"/>	Origem, confiabilidade

2	Verificar estrutura dos dados (colunas, linhas, categorias)	<input type="checkbox"/>	Dados bem distribuídos?
3	Detectar e remover valores nulos	<input type="checkbox"/>	Informar se foram excluídos ou substituídos
4	Identificar e remover duplicatas	<input type="checkbox"/>	Duplicações de registros ou campos
5	Corrigir formatação de dados (datas, números, textos)	<input type="checkbox"/>	Padronizar formatos
6	Corrigir erros ortográficos ou inconsistências	<input type="checkbox"/>	Ex: "informatica" e "Informática"
7	Padronizar nomes e categorias	<input type="checkbox"/>	Ex: Nome completo, gênero, estado civil
8	Validar valores fora do padrão (ex: idade > 120)	<input type="checkbox"/>	Verificar e corrigir outliers
9	Criar tabelas lógicas separadas (Clientes, Vendas etc.)	<input type="checkbox"/>	Usar normalização básica
10	Documentar campos (dicionário de dados)	<input type="checkbox"/>	Nome, tipo, descrição
11	Garantir unicidade de chaves (ID, CPF etc.)	<input type="checkbox"/>	Verificar se não há repetições
12	Salvar versão final organizada dos dados	<input type="checkbox"/>	Nome do arquivo: dados_tratados_final.xlsx

### Modelo de Dicionário de Dados para MySQL

Tabela	Campo	Tipo de Dados	Descrição	Chave Primária	Chave Estrangeira	Nulo	Padrão	Comentários/Restrições
<b>paciente</b>	id_paciente	INT(11)	Identificador único do paciente	Sim	Não	Não	AUTO_INCREMENT	
<b>paciente</b>	nome	VARCHAR(100)	Nome completo do	Não	Não	Não	NULL	

			paciente					
paciente	cpf	VARCHAR(14)	CPF do paciente	Não	Não	Não	NULL	Formato XXX.XXX.XXX-XX, único
paciente	data_nascimento	DATE	Data de nascimento	Não	Não	Sim	NULL	
paciente	telefone	VARCHAR(20)	Número de telefone	Não	Não	Sim	NULL	
paciente	email	VARCHAR(100)	Email do paciente	Não	Não	Sim	NULL	