# Transformando dados

## O QUE É TRANSFORMAÇÃO DE DADOS?

Uma mulher apresentando dados, uma mão segurando uma medalha, duas pessoas conversando, um leme de navio sendo dirigido, duas pessoas cumprimentando-se
Nesta leitura, você explorará como os dados são transformados e as diferenças entre dados extensos e longos. A **transformação de dados** é o processo de alterar o formato, a estrutura ou os valores dos dados. Como analista de dados, há uma boa chance de você precisar transformar os dados em algum ponto para facilitar sua análise.

A transformação de dados geralmente envolve:

- **Adicionar, copiar ou replicar dados**
- **Excluindo campos ou registros**
- **Padronizando os nomes das variáveis**
- **Renomear, mover ou combinar colunas em um banco de dados**
- **Unindo um conjunto de dados com outro**
- **Salvando um arquivo em um formato diferente. Por exemplo, salvar uma planilha como um arquivo de valores separados por vírgula (CSV).**

## POR QUE TRANSFORMAR DADOS?

As metas para a transformação de dados podem ser:

- **Organização** de dados: dados melhor organizados são mais fáceis de usar
- **Compatibilidade** de dados: diferentes aplicativos ou sistemas podem usar os mesmos dados
- **Migração** de dados: dados com formatos correspondentes podem ser movidos de um sistema para outro
- **Mesclagem** de dados: dados com a mesma organização podem ser mesclados
- **Aprimoramento** de dados: os dados podem ser exibidos com campos mais detalhados
- **Comparação** de dados: comparações de igual para igual dos dados podem então ser feitas

### EXEMPLO DE TRANSFORMAÇÃO DE DADOS: MESCLAGEM DE DADOS

Mario é um encanador dono de uma empresa de encanamento. Depois de anos no negócio, ele compra outra empresa de encanamento. Mario deseja mesclar as informações do cliente de sua empresa recém-adquirida com as suas, mas a outra empresa usa um banco de dados diferente. Então, Mario precisa tornar os dados compatíveis. Para isso, ele precisa transformar o formato dos dados da empresa adquirida. Em seguida, ele deve remover as linhas duplicadas dos clientes que eles tinham em comum. Quando os dados são compatíveis e juntos, a empresa de encanamento de Mario terá um banco de dados de clientes completo e mesclado.

### EXEMPLO DE TRANSFORMAÇÃO DE DADOS: ORGANIZAÇÃO DE DADOS (LONGO PARA AMPLO)

Para facilitar a criação de gráficos, você também pode precisar transformar dados longos em dados extensos. Considere o seguinte exemplo de transformação dos preços das ações (coletados como dados longos) em dados amplos.
**Dados longos** são dados em que **cada linha contém um único ponto de dados** para um determinado item. No exemplo de dados longos abaixo, preços de ações individuais (pontos de dados) foram coletados para Apple (AAPL), Amazon (AMZN) e Google (GOOGL) (itens específicos) nas datas fornecidas.

**EXEMPLO DE DADOS LONGOS: PREÇOS DAS AÇÕES**

| Symbol | Date | Open |
|---|---|---|
| AAPL | 2018-09-18 | 217.79 |
| AAPL | 2018-09-17 | 222.15 |
| AAPL | 2018-09-14 | 225.75 |
| AAPL | 2018-09-13 | 223.52 |
| AMZN | 2018-09-18 | 1918.65 |
| AMZN | 2018-09-17 | 1954.73 |
| AMZN | 2018-09-14 | 1992.93 |
| AMZN | 2018-09-13 | 2000 |
| GOOGL | 2018-09-18 | 1162.66 |
| GOOGL | 2018-09-17 | 1177.77 |
| GOOGL | 2018-09-14 | 1188 |
| GOOGL | 2018-09-13 | 1179.7 |

**Dados amplos** são dados em que **cada linha contém vários pontos de dados** para os itens específicos identificados nas colunas.

**EXEMPLO DE DADOS ABRANGENTES: PREÇOS DAS AÇÕES**

| Symbol | AAPL | AMZN | GOOGL |
|---|---|---|---|
| Date | | | |
| 2018-09-13 | 223.52 | 2000 | 1179.7 |
| 2018-09-14 | 225.75 | 1992.93 | 1188 |
| 2018-09-17 | 222.15 | 1954.73 | 1177.77 |
| 2018-09-18 | 217.79 | 1918.65 | 1162.66 |

Com os dados transformados em dados amplos, você pode criar um gráfico comparando como as ações de cada empresa mudaram no mesmo período de tempo. Você pode notar que todos os dados incluídos no formato longo também estão no formato largo. Mas dados amplos são mais fáceis de ler e entender. É por isso que os analistas de dados normalmente transformam dados longos em dados extensos com mais frequência do que transformam dados extensos em dados longos. A tabela a seguir resume quando cada formato é preferido:

| **Dados amplos são preferidos quando** | **Dados longos são preferidos quando** |
| --- | --- |
| Criação de tabelas e gráficos com algumas variáveis sobre cada assunto | Armazenando muitas variáveis sobre cada assunto, por exemplo, 60 anos de taxas de juros para cada banco |
| Comparando gráficos de linha simples | Execução de análises estatísticas avançadas ou gráficos |