

## Ciladas comuns da limpeza de dados

Nesta leitura, você aprenderá a importância da limpeza de dados e como identificar erros comuns. Erros comuns a evitar

- **Não verificação de erros ortográficos:** Os erros ortográficos podem ser tão simples como erros de digitação ou de entrada. Na maioria das vezes a ortografia errada ou erros gramaticais comuns podem ser detectados, mas fica mais difícil com coisas como nomes ou endereços. Por exemplo, se você estiver trabalhando com uma tabela de planilhas de dados do cliente, você pode encontrar um cliente chamado "John" cujo nome foi inserido incorretamente como "Jon" em alguns lugares. A verificação ortográfica da planilha provavelmente não marcará isto, portanto, se você não verificar novamente os erros ortográficos e apanhar isto, sua análise terá erros nela.
- **Esquecer de documentar erros:** Documentar seus erros pode ser uma grande economia de tempo, pois ajuda a evitar esses erros no futuro, mostrando-lhe como resolvê-los. Por exemplo, você pode encontrar um erro em uma fórmula em sua planilha. Você descobre que algumas das datas em uma de suas colunas não foram formatadas corretamente. Se você tomar nota desta correção, poderá referenciá-la na próxima vez que sua fórmula for quebrada, e obter um avanço na solução de problemas. Documentar seus erros também ajuda a acompanhar as mudanças em seu trabalho, para que você possa recuar se uma correção não funcionou.
- **Não verificação de valores de campos errados:** Um valor errado acontece quando os valores são inseridos no campo errado. Estes valores ainda podem ser formatados corretamente, o que os torna mais difíceis de serem capturados se você não for cuidadoso. Por exemplo, você pode ter um conjunto de dados com colunas para cidades e países. Estes são o mesmo tipo de dados, por isso são fáceis de misturar. Mas se você estivesse tentando encontrar todas as instâncias da Espanha na coluna do país, e a Espanha tivesse sido inserida erroneamente na coluna da cidade, você perderia pontos de dados-chave. Garantir que seus dados tenham sido inseridos corretamente é a chave para uma análise precisa e completa.
- **Ignorando os valores ausentes:** Valores em falta em seu conjunto de dados podem criar erros e dar-lhe conclusões imprecisas. Por exemplo, se você estivesse tentando obter o número total de vendas dos últimos três meses, mas faltasse uma semana de transações, seus cálculos não seriam imprecisos. Como melhor prática, tente manter seus dados o mais limpos possível, mantendo a completude e consistência.
- **Apenas olhando para um subconjunto dos dados:** É importante pensar em todos os dados relevantes quando você estiver limpando. Isto ajuda a entender toda a história que os dados estão contando, e que você está prestando atenção a todos os possíveis erros. Por exemplo, se você estiver trabalhando com dados sobre padrões de migração de aves de diferentes fontes, mas você limpa apenas uma fonte, você pode não perceber que alguns dos dados estão sendo repetidos. Isto causará problemas em sua análise mais tarde. Se você quiser evitar erros comuns como duplicatas, cada campo de seus dados requer a mesma atenção.

- **Perder a noção dos objetivos comerciais:** Quando você estiver limpando dados, você pode fazer novas e interessantes descobertas sobre seu conjunto de dados - mas você não quer que essas descobertas o distraiam da tarefa em mãos. Por exemplo, se você estava trabalhando com dados meteorológicos para encontrar o número médio de dias chuvosos em sua cidade, você também pode notar alguns padrões interessantes sobre a queda de neve. Isso é realmente interessante, mas não está relacionado à pergunta que você está tentando responder no momento. Ser curioso é ótimo! Mas tente não deixar que isso o distraia da tarefa em mãos.
- **Não consertando a fonte do erro:** A correção do erro em si é importante. Mas se esse erro é realmente parte de um problema maior, você precisa encontrar a origem do problema. Caso contrário, você terá que continuar corrigindo esse mesmo erro uma e outra vez. Por exemplo, imagine que você tenha uma planilha de equipe que acompanhe o progresso de todos. A tabela continua quebrando porque pessoas diferentes estão inserindo valores diferentes. Você pode continuar resolvendo todos esses problemas um a um, ou pode configurar sua tabela para agilizar a entrada de dados de modo que todos estejam na mesma página. A abordagem da fonte dos erros em seus dados lhe poupará muito tempo a longo prazo.
- **Não analisar o sistema antes da limpeza dos dados:** Se quisermos limpar nossos dados e evitar erros futuros, precisamos entender a causa raiz de seus dados sujos. Imagine que você é um mecânico de automóveis. Você encontraria a causa do problema antes de começar a consertar o carro, certo? O mesmo vale para os dados. Primeiro, você descobre de onde vêm os erros. Talvez seja devido a um erro de entrada de dados, não ter sido feita uma verificação ortográfica, falta de formatos, ou de duplicatas. Então, quando você entender de onde vêm os dados ruins, você pode controlá-los e manter seus dados limpos.
- **Não fazer backup de seus dados antes da limpeza dos dados:** É sempre bom ser pró-ativo e criar seu backup de dados antes de iniciar a limpeza de seus dados. Se seu programa falhar, ou se suas alterações causarem um problema em seu conjunto de dados, você pode sempre voltar à versão salva e restaurá-la. O simples procedimento de backup de seus dados pode lhe poupar horas de trabalho - e o mais importante, uma dor de cabeça.
- **Não contabilizando a limpeza de dados em seus prazos/processos:** Todas as coisas boas levam tempo, e isso inclui a limpeza de dados. É importante ter isso em mente ao passar por seu processo e ao olhar para seus prazos. Quando você reserva tempo para a limpeza dos dados, isso ajuda a obter uma estimativa mais precisa dos ETAs para as partes interessadas, e pode ajudá-lo a saber quando solicitar um ETA ajustado.

## Recursos adicionais

Consulte estas listas "top ten" para limpeza de dados no Microsoft Excel e Google Sheets para ajudá-lo a evitar os erros mais comuns:

- [As dez melhores maneiras de limpar seus dados](#)  
: reveja um guia ordenado de limpeza de dados no Microsoft Excel.
- [10 dicas do Google Workspace para limpar dados](#)  
: aprenda as melhores práticas de limpeza de dados no Google Sheets.