

Dados sujos (Dirty data) 🖌️

Alguns tipos de dados sujos são:

- **Dados duplicados (Duplicate data):** qualquer dado que aparece mais de uma vez
- **Dados desatualizados (Outdated data):** dados que deveriam ser substituídos por uma versão mais recente
- **Dados incompletos (Incomplete data):** dados com informações importantes faltando
- **Dados incorretos (Inaccurate data):** dados completos porém incorretos
- **Dados inconsistentes (Inconsistent data):** qualquer dado que use diferentes formas de representar a mesma coisa

Limpando dados 🧼

Ao limpar os dados, certifique-se de:

1. **Checar erros de gramática**
 2. **Documentar como corrigir os erros que encontrar:** com isso é possível agilizar o processo de limpeza na próxima vez
 3. **Checar dados em colunas erradas (misfielded values):** pode acontecer de um dado correto ter sido colocado na coluna errada, por exemplo um nome de um país ter sido colocado na coluna de cidade e vice-versa.
 4. **Não ignorar valores nulos:** caso contrário, essa prática pode resultar em análises incorretas ou incompletas
 5. **Limpar os dados como um todo:** olhar apenas um pedaço dos dados, ou apenas alguns campos pode levar a informações repetidas e outros problema
 6. **Mantenha o foco no objetivo:** a limpeza deve ser focada no objetivo final da análise
 7. **Corrija a raiz do problema:** ao encontrar um problema, além de corrigi-lo, é importante corrigir o que causou ele, para não perder mais tempo no futuro com o mesmo erro
 8. **Analisar o sistema antes de começar a limpeza:** é importante entender da onde aquele erro veio e como foi causado, para saber como corrigir corretamente.
 9. **Sempre fazer um Backup antes da limpeza**
 10. **Considerar o tempo de limpeza no prazo final:** o processo de limpeza toma tempo, então é sempre importante considerar ele para se manter dentro dos prazos.
- [Limpando os dados no Excel](#)
 - [Limpando do dados no Google Sheets](#)

Automatizando a limpeza de dados

- [Automatizando análise de dados](#)
- [Automatizando Big-Data](#)
- [10 ferramentas para automação](#)