

Como verificar a limpeza de dados: Uma lista de verificação

Nesse artigo, você encontrará uma lista de verificação dos problemas comuns; consulte-a durante a verificação da limpeza de dados, independentemente da ferramenta que usar. Quando se trata da verificação da limpeza de dados, não há uma abordagem ou lista de verificação únicas que possam ser aplicadas universalmente a todos os projetos. Cada projeto possui seus próprios requisitos de organização e dados que levam a uma lista única de coisas que você deve conferir durante a verificação.



Lembre-se: conforme for obtendo mais dados ou entender melhor as metas do projeto, você poderá querer voltar a algumas ou todas essas etapas.

CORRIGIR OS PROBLEMAS MAIS COMUNS

Certifique-se de ter identificado os problemas mais comuns e corrija-os, incluindo:

- **Origem dos erros.** Você usou as ferramentas e funções certas para encontrar a origem dos erros no seu conjunto de dados?
- **Dados nulos.** Você procurou por NULOS com filtros e formatação condicional?
- **Palavras digitadas incorretamente.** Você localizou todas as palavras com erro de digitação?
- **Números digitados incorretamente.** Você verificou se os dados numéricos foram digitados corretamente?
- **Caracteres e espaços extras.** Você excluiu os caracteres ou espaços extras com a função TRIM?
- **Duplicatas.** Você excluiu as duplicatas nas planilhas ou SQL com as funções Remove duplicates ou DISTINCT, respectivamente?
- **Tipos de dados incompatíveis.** Você verificou se os dados numéricos, de datas e strings foram convertidos corretamente?
- **Strings desorganizadas (inconsistentes).** Você verificou se todas as strings são consistentes e pertinentes?

- **Formatos de dados desorganizados (inconsistentes).** Você formatou as datas de forma consistente no conjunto de dados?
- **Identificações (colunas) variáveis incorretas.** Você nomeou suas colunas de forma adequada?
- **Dados truncados.** Você verificou se há dados ausentes ou truncados que exigem correção?
- **Lógica nos negócios.** Com base em seu conhecimento nos negócios, você verificou se os dados são coerentes?

REVISAR A META DO SEU PROJETO

Depois de concluídas as tarefas de limpeza dos dados, revise a meta do seu projeto e verifique se seus dados continuam alinhados à meta. Esse é um processo contínuo que deverá ser feito durante todo o projeto; há, no entanto, três etapas para levar em consideração ao refletir a respeito:

- Confirmar o problema dos negócios
- Confirmar a meta do projeto
- Verificar se os dados têm potencial para solucionar o problema e estão alinhados à meta