



**great  
expectations**

**para Python**

**Ferramenta de Data Quality**

# Importância do Data Quality



Data Quality refere-se a **quão precisos, completos, consistentes, válidos e atualizados** são os dados utilizados em um processo.

## Por que é importante?

- ✓ tomar decisões mais seguras
- 🔍 facilitar o processo de ETL
- 🔄 tornar os processos mais eficientes
- 💡 criar análises e relatórios mais confiáveis
- 🤖 evitar erros em sistemas e automações
- 📉 reduzir custos com retrabalho
- 🔍 garantir conformidade com leis (LGPD)

## Step no pipeline

Antes da carga (ETL/ELT)

Durante o processamento em tempo real

Como etapa de validação em projetos

**dados brutos em ETL → validação (Great Expectations) → análise**

# Great Expectations: conhecendo a ferramenta



## O que é

Great Expectations é uma **ferramenta open source** criada pela empresa Superconductive para ajudar equipes a **garantir a qualidade dos dados em pipelines de dados e análises**.



## O que faz

Permite **criar, executar e documentar testes automatizados** (chamados de **expectations**) que definem o que é considerado um "dado válido" no processamento em questão.



## Potencial

É ideal para validar dados em ETL/ELT, Machine Learning, BI e Data Warehouses, **reduzindo erros e aumentando a confiança nas análises**, inclusive as feitas com dados em tempo real.



# Projeto: Sleep efficiency - Kaggle + Google Colab



**greatexpectations.io**

**[https://www.youtube.com/watch?v=7Nk0HiiWi\\_Q](https://www.youtube.com/watch?v=7Nk0HiiWi_Q)**