

A estrutura de dados

Os dados estão em todos os lugares e podem ser armazenados de várias maneiras. Duas categorias gerais de dados são:

- **Dados estruturados:** organizado em um determinado formato, como linhas e colunas.
- **Dados não estruturados:** não organizado de forma fácil de identificar.

Por exemplo, ao avaliar seu restaurante favorito online, você está criando dados estruturados. Mas quando você usa o Google Earth para verificar uma imagem de satélite da localização de um restaurante, está usando dados não estruturados. Aqui está uma atualização sobre as características de dados estruturados e não estruturados:

Dados Estruturados	Dados Não Estruturados
Tipos de dados definidos	Tipos de dados variados
Na maioria das vezes, dados quantitativos	Na maioria das vezes, dados qualitativos
Fácil de organizar	Difícil de pesquisar
Fácil de pesquisar	Oferece mais liberdade para análise
Fácil de analisar	Armazenado em data lakes, data warehouses e bancos de dados NoSQL
Armazenado em bancos de dados relacionais e data warehouses	Não pode ser colocado em linhas e colunas
Contido em linhas e colunas	Exemplos: mensagens de texto, comentários de mídia social, transcrições de chamadas telefônicas, vários arquivos de registro, imagens, áudio, vídeo
Exemplos: Excel, Planilhas Google, SQL, dados do cliente, registros telefônicos, histórico de transações	

DADOS ESTRUTURADOS

Conforme descrito anteriormente, os **dados estruturados** são organizados em um determinado formato. Isso torna mais fácil armazenar e consultar as necessidades de negócios. Se os dados são exportados, a estrutura acompanha os dados.

DADOS NÃO ESTRUTURADOS

Os **dados não estruturados** não podem ser organizados de nenhuma maneira facilmente identificável. E há muito mais dados não estruturados do que estruturados no mundo. Arquivos de vídeo e áudio, arquivos de texto, conteúdo de mídia social, imagens de satélite, apresentações, arquivos PDF, respostas de pesquisas abertas e sites da Web são todos qualificados como tipos de dados não estruturados.

A QUESTÃO DA JUSTIÇA

A falta de estrutura torna os dados não estruturados difíceis de pesquisar, gerenciar e analisar. Porém os avanços recentes em inteligência artificial e algoritmos de aprendizado de máquina estão começando a mudar isso. Agora, o novo desafio que os cientistas de dados enfrentam é garantir que essas ferramentas sejam inclusivas e imparciais. Caso contrário, certos elementos de um conjunto de dados serão mais ponderados e / ou representados do que outros. E, de acordo com o que você está aprendendo, um conjunto de dados injusto não representa com precisão a população, causando resultados distorcidos, baixos níveis de precisão e análises não confiáveis.