

## O que fazer quando não há dados

Anteriormente, você aprendeu como ainda pode fazer uma análise usando dados de proxy se não tiver dados. Você pode ter algumas dúvidas sobre dados de proxy, portanto, esta leitura fornecerá mais alguns exemplos dos tipos de conjuntos de dados que podem servir como fontes de dados alternativas.

### EXEMPLOS DE DADOS DE PROXY

Às vezes, os dados para apoiar um objetivo de negócios não estão prontamente disponíveis. É quando os dados de proxy são úteis. Dê uma olhada nos seguintes cenários e onde os dados de proxy entram para cada exemplo:

Cenário de negócios	Como os dados de proxy podem ser usados
Um novo modelo de carro foi lançado há alguns dias e a concessionária mal pode esperar até o final do mês para que os dados de vendas cheguem. Eles querem projeções de vendas agora.	O analista aproxima o número de cliques nas especificações do carro no site da concessionária como uma estimativa de vendas potenciais na concessionária.
Um novo produto de carne à base de plantas foi recentemente estocado em mercearias e o fornecedor precisa estimar a demanda nos próximos quatro anos.	O analista procura os dados de vendas de um substituto de peru feito de tofu que está no mercado há vários anos.
A Câmara de Comércio quer saber como uma campanha de turismo afetará as viagens para sua cidade, mas os resultados da campanha ainda não estão disponíveis publicamente.	O analista procura os dados históricos de reservas aéreas para a cidade um a três meses depois que uma campanha semelhante foi realizada seis meses antes.

### CONJUNTOS DE DADOS PÚBLICOS EXAMINADOS

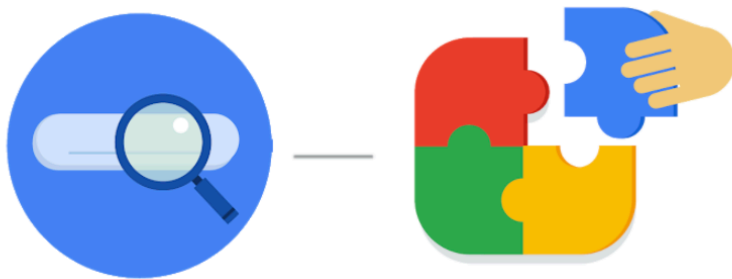
Se você faz parte de uma grande organização, pode ter acesso a muitas fontes de dados. Mas se você estiver procurando por algo específico ou um pouco fora de sua linha de negócios, também poderá usar conjuntos de dados abertos ou públicos. (Você pode consultar este artigo da [Towards Data Science](#)

para uma breve explicação da diferença entre dados abertos e públicos.)

Aqui está um exemplo. Uma versão nasal de uma vacina foi recentemente disponibilizada. Uma clínica quer saber o que esperar das contraindicações, mas acabou de começar a coletar dados primários de seus pacientes. Uma **contraindicação** é uma condição que pode fazer com que um paciente não tome uma vacina devido aos danos que ela causaria se fosse tomada. Para estimar o número de possíveis contraindicações, um analista de dados aproxima um conjunto de dados aberto de um teste da versão de injeção da vacina. O analista seleciona um subconjunto de dados com perfis de pacientes que mais se aproximam da composição dos pacientes na clínica.

Há muitas maneiras de compartilhar e colaborar em dados dentro de uma comunidade. A Kaggle ([kaggle.com](https://www.kaggle.com)

), que apresentamos anteriormente, possui conjuntos de dados em vários formatos, incluindo o tipo mais básico, arquivos com valores separados por vírgula (CSV).



CSV, JSON, conjuntos de dados SON, SQLite e BigQuery

- CSV: Confira este conjunto de dados de [clientes de cartão de crédito](#), que tem informações de 10.000 clientes, incluindo idade, salário, estado civil, limite do cartão de crédito, categoria do cartão de crédito etc. (CC0: Domínio Público, Sakshi Goyal).
- JSON: Confira este conjunto de dados do JSON para [vídeos de tendências do YouTube](#) (CC0: Domínio Público, Mitchell J).
- SQLite: Confira este conjunto de dados da SQLite para 24 anos de [dados de incêndios florestais nos EUA](#) (CC0: Domínio Público, Rachael Tatman).
- BigQuery: Confira este conjunto de dados de amostra do [Google Analytics 360](#) da Google Merchandise Store (CC0 Public Domain, Google BigQuery).

Consulte a [documentação para conjuntos](#)

do Kaggle de dados para obter mais informações e pesquise e explore conjuntos de dados por conta própria em [kaggle.com/datasets](https://www.kaggle.com/datasets)

Tal como acontece com todos os outros tipos de conjuntos de dados, fique atento a dados duplicados e 'Nulo' em conjuntos de dados abertos. Nulo na maioria das vezes significa

que um campo de dados não foi atribuído (deixado em branco), mas às vezes Nulo pode ser interpretado como o valor 0. É importante entender como o Nulo foi usado antes de começar a analisar um conjunto de dados com dados Nulo.