

O que fazer ao encontrar um problema com os dados

Quando você está se preparando para a análise de dados, pode perceber que não tem os dados de que precisa ou não tem o suficiente. Em alguns casos, você pode usar o que é conhecido como dados de proxy no lugar dos dados reais. Pense nisso como substituir óleo por manteiga em uma receita quando você não tem manteiga. Em outros casos, não há substituto razoável e sua única opção é coletar mais dados.

Considere os seguintes problemas de dados e sugestões sobre como contorná-los.

PROBLEMA DE DADOS 1: FALTA DE DADOS

Soluções possíveis,	Exemplos de soluções na vida real
Colete os dados em pequena escala para realizar uma análise preliminar e, em seguida, solicite	Se você estiver pesquisando funcionários sobre o que eles pensam sobre um novo plano de desempenho e bônus, use uma amostra para uma
Se não houver tempo para coletar dados, realize a análise usando dados de proxy de outros	Se você estiver analisando os horários de pico de viagens para usuários de transporte público, mas não tiver os dados de uma cidade específica, use os

PROBLEMA DE DADOS 2: MUITO POUCOS DADOS

Soluções possíveis,	Exemplos de soluções na vida real
Faça a análise usando dados de proxy junto com dados reais.	Se você estiver analisando tendências para donos de golden retrievers, aumente seu conjunto de dados incluindo os dados de donos de labradores.
Ajuste sua análise para alinhar com os dados que você já possui.	Se faltarem dados para jovens de 18 a 24 anos, faça a análise, mas observe a seguinte limitação em seu relatório: <i>esta conclusão se aplica apenas a adultos com 25 anos ou mais.</i>

PROBLEMA DE DADOS 3: DADOS ERRADOS, INCLUINDO DADOS COM ERROS*

Soluções possíveis,	Exemplos de soluções na vida real
Se você tiver os dados errados porque os requisitos foram mal compreendidos, comunique os requisitos novamente.	Se você precisa dos dados de eleitores do sexo feminino e recebeu os dados de eleitores do sexo masculino, reafirme suas necessidades.
Identifique erros nos dados e, se possível, corrija-os na origem, procurando um padrão nos erros.	Se seus dados estiverem em uma planilha e houver uma declaração condicional ou booleana fazendo com que os cálculos estejam errados, altere a declaração condicional em vez de apenas corrigir os valores calculados.
Se você não puder corrigir os erros de dados sozinho, poderá ignorar os dados errados e prosseguir com a análise se o tamanho da amostra ainda for grande o suficiente e ignorar os dados não causará viés sistemático.	Se seu conjunto de dados foi traduzido de um idioma diferente e algumas das traduções não fazem sentido, ignore os dados com tradução ruim e prossiga com a análise dos outros dados.

*** Observação importante: às vezes, dados com erros podem ser um sinal de alerta de que os dados não são confiáveis. Use seu melhor julgamento.*

Use a seguinte árvore de decisão como um lembrete sobre como lidar com erros de dados ou dados insuficientes:

Data Errors

