

Limitações dos dados

Os dados são poderosos, mas têm suas limitações. A opinião pessoal de alguém conseguiu influenciar os números? Os dados estão mostrando o panorama completo? Parte de ser um bom analista de dados é saber quais são os limites dos dados e se antecipar a eles. Este texto explica como fazer isso.



The case of incomplete (or nonexistent!) data

O caso dos dados incompletos (ou inexistentes!)

Se tiver dados incompletos ou inexistentes, você poderá perceber durante a análise que não conta com dados suficientes para chegar a uma conclusão. Ou você pode estar resolvendo um problema completamente diferente! Por exemplo, suponhamos que você está procurando funcionários que receberam um certificado específico, mas descobre que os registros de certificado abrangem somente os últimos dois anos da empresa. Você poderá usar esses dados, mas precisará esclarecer as limitações dessa análise. Talvez você consiga encontrar uma fonte de dados alternativa se entrar em contato com a empresa que forneceu o capacitação. Mas, para ter certeza, você precisa informar claramente que tem um conjunto de dados incompleto até que outros dados estejam disponíveis.



Don't miss misaligned data

É preciso alinhar os dados

Se estiver coletando dados de outra equipe e usando planilhas existentes, é bom lembrar que as pessoas usam diferentes regras de negócios. Portanto, uma equipe pode definir e medir parâmetros de maneira completamente diferente de outra. Por exemplo, se uma métrica é o número total de estagiários em um programa de certificação, uma equipe pode registrar todas as pessoas inscritas no programa e outra equipe pode contar somente os que concluíram o programa. Nesses casos, estabelecer como os parâmetros serão medidos logo no início padroniza os dados em toda a empresa, gerando maior confiabilidade e acurácia. Isso fará com que as comparações entre as equipes sejam significativa e gerem muitos insights.



Deal with dirty data

Lidar com dados sujos

A expressão “dados sujos” refere-se a dados que contêm erros. Os dados sujos podem gerar uma queda na produtividade, gastos desnecessários e a tomada de decisões incorretas. Uma boa limpeza de dados pode ajudar a evitar tudo isso. Lembramos que a limpeza de dados é o processo de corrigir ou remover dados incorretos, corrompidos, com formato incorreto, duplicados ou incompletos no conjunto de dados. Ao encontrar e corrigir os erros, mantendo um registro das alterações feitas, você pode evitar um desastre de dados. Você aprenderá como limpar dados mais adiante neste treinamento.



Tell a clear story

Contar uma história transparente

Avinash Kaushik, *Digital Marketing Evangelist* do Google, tem várias dicas ótimas para analistas de dados no [blog: Navalha de Occam](#)

. Veja a seguir algumas das práticas recomendadas por ele para contar boas histórias usando dados.

- **Compare os mesmos tipos de dados:** Os dados podem se misturar ao criar gráficos para visualização. Compare sempre os mesmos tipos de dados e verifique que nenhum segmento do gráfico exiba métricas diferentes.
- **Visualize com cuidado:** Uma queda de 0,01% em uma pontuação pode ser enorme quando examinada de perto. Para que seu público-alvo entenda o panorama completo, é recomendável definir o eixo Y como 0.
- **Não inclua gráficos desnecessários:** Se uma tabela for capaz de mostrar os dados rapidamente, fique com a tabela e esqueça os gráficos de pizza ou de barras. Seu público-alvo agradecerá a clareza.
- **Teste a significância estatística:** Às vezes, dois conjuntos de dados parecerão diferentes, mas é preciso testar se essa diferença é real e importante. Faça testes estatísticos para ver se você pode confiar nessa diferença.

- **Preste atenção ao tamanho da amostra:** Colete muitos dados. Se a amostra for pequena, algumas respostas incomuns podem distorcer os resultados. Se perceber que não tem dados suficientes, tome cuidado ao usá-los para criar conceitos. Procure oportunidades para coletar mais dados e registre essas tendências em períodos mais longos.



Be the judge

Crie conceitos

Em qualquer organização, boa parte do trabalho do analista de dados consiste em criar conceitos. Ao conhecer as limitações dos dados, você pode criar conceitos que ajudem as pessoas a tomarem decisões baseadas em dados. Os dados são uma ferramenta extremamente poderosa para tomar decisões, mas se estiverem incompletos, mal alinhados ou sujos, podem levar a conclusões errôneas. Tome as medidas necessárias para se certificar de que os dados estejam completos e sejam consistentes. Limpe os dados antes de começar a análise para economizar seu tempo e esforço e, possivelmente, os de outras pessoas.