

Inspecionando um conjunto de dados: Um tour prático guiado

Como analista de dados, você usará os dados para responder perguntas e resolver problemas. Quando você analisa dados e tira conclusões, você está gerando insights que podem influenciar as decisões comerciais, impulsionar mudanças positivas e ajudar suas partes interessadas a atingir seus objetivos.

Antes de iniciar uma análise, é importante inspecionar seus dados para determinar se eles contêm as informações específicas que você precisa para responder às perguntas de suas partes interessadas. Em qualquer dado conjunto de dados, pode ser o caso que:

- Os dados não estão lá (você tem dados de sanduíche, mas precisa de dados de pizza)
- Os dados são insuficientes (você tem dados de pizza para 1-7 de junho, mas precisa de dados para o mês inteiro de junho)
- Os dados estão incorretos (seus dados de pizza listam o custo de uma fatia como \$250, o que o faz questionar a validade do conjunto de dados)

A inspeção de seu conjunto de dados o ajudará a identificar quais perguntas são respondidas e quais dados ainda estão faltando. Você pode ser capaz de recuperar esses dados de uma fonte externa ou pelo menos recomendar às partes interessadas que outra fonte de dados seja utilizada.

Nesta leitura, imagine que você seja um analista de dados inspecionando dados de planilhas para determinar se é possível responder às perguntas de suas partes interessadas.

CENÁRIO:

Você é um analista de dados que trabalha para uma empresa de sorvetes. A administração está interessada em melhorar as vendas de sorvetes da empresa.

A empresa tem coletado dados sobre suas vendas - mas não muito. Os dados disponíveis são de uma fonte de dados interna e se baseiam nas vendas para 2019. Foi solicitado que você revisse os dados e fornecesse algumas informações sobre as vendas de sorvetes da empresa. De preferência, a administração gostaria de respostas para as seguintes perguntas:

1. Qual é o sabor mais popular de sorvete?
2. Como a temperatura afeta as vendas?
3. Como os fins de semana e feriados afetam as vendas?
4. Qual é a diferença entre rentabilidade para clientes novos e clientes que retornam?

INSPECIONE OS DADOS

PERGUNTA 1: QUAL É O SABOR MAIS POPULAR DE SORVETE?

Para descobrir o sabor mais popular, é preciso primeiro definir o que se entende por "popular". O sabor mais popular é aquele que gerou mais receitas em 2019? Ou é o sabor que teve o maior número de unidades vendidas em 2019? Às vezes suas escolhas de medição são limitadas pelos dados que você tem - você pode rever sua planilha para descobrir se qualquer uma destas definições de "popular" faz sentido com base nos dados disponíveis.

Clique na guia **sabores** em sua planilha para visualizar os dados relevantes.

A planilha de **sabores** tem três colunas e 209 linhas de dados. Os cabeçalhos das colunas são **semana**, **unidades vendidas** e **sabor**. Este conjunto de dados não veio com uma descrição dos dados, então você tem que descobrir o significado das colunas por conta própria. Com base nos dados, você deduz que estas colunas fornecem informações sobre o número de unidades vendidas para cada sabor de sorvete, por semana, em 2019

	A	B	C
1	week	units sold	flavor
2	1	6	chocolate
3	1	15	lemon
4	1	12	strawberry
5	1	6	vanilla
6	2	16	chocolate
7	2	7	lemon
8	2	7	strawberry
9	2	18	vanilla
10	3	5	chocolate

Neste caso, você pode descobrir qual é o sabor mais popular usando unidades vendidas como sua medida. Em particular, pode-se usar a coluna de **unidades vendidas** para calcular o número total de unidades vendidas durante o ano para cada sabor. Infelizmente, o conjunto de dados não fornece a quantidade de vendas anuais por sabor. Neste caso, seu próximo passo seria perguntar a suas partes interessadas se os dados das vendas anuais por sabor estão disponíveis em outra fonte. Caso contrário, você pode acrescentar uma declaração sobre as limitações dos dados atuais à sua análise.

PERGUNTA 2: COMO A TEMPERATURA AFETA AS VENDAS?

Para explorar sua segunda pergunta, clique na aba **temperaturas** e confira os dados. A planilha de **temperaturas** tem duas colunas e 366 linhas de dados. Os cabeçalhos das colunas são **temperatura** e **vendas**. Os dados podem mostrar o total de vendas por temperatura em 2019 (por exemplo, a primeira entrada pode somar \$36,69 em vendas por três dias separados que cada um teve uma alta de 60 graus). Ou, os dados podem mostrar um instantâneo das vendas e temperatura para cada dia em 2019 (por exemplo, a primeira entrada pode se referir a um único dia com uma alta de 60 graus e \$39,69 em vendas).

	A	B
1	temperature	sales
2	60	39.69
3	80	61.59
4	58	33.44
5	96	80.02
6	95	80.75
7	54	31.94
8	74	53.62
9	58	36.79
10	54	28.37

Então, o que é isso? É provavelmente um retrato diário porque há 365 entradas para temperatura, e várias linhas com a mesma temperatura e valores de vendas diferentes. Isso implica que cada entrada é para um único dia e não um resumo de vários dias. Entretanto, sem mais informações, você não pode ter certeza. Além disso, você não sabe se os dados atuais estão listados em ordem consecutiva por data ou em uma ordem diferente. Seu próximo passo seria entrar em contato com o proprietário do conjunto de dados para esclarecimentos.

Se for verificado que a temperatura afeta as vendas, você será capaz de oferecer a suas parte interessadas uma visão como a que se segue: "Quando as máximas diárias estão acima de X graus, a média de vendas de sorvete aumenta em Y. Portanto, a empresa deve planejar o aumento do estoque durante esses tempos para maximizar as vendas".

PERGUNTA 3: COMO OS FINS DE SEMANA E FERIADOS AFETAM AS VENDAS?

A seguir, clique na guia de **vendas** para visualizar os dados sobre as datas de venda. A **planilha** de vendas tem duas colunas e 366 linhas de dados. Os cabeçalhos das colunas são **data** e **vendas**. Esses dados são, muito provavelmente, o total de vendas diárias em 2019, já que as vendas são registradas para cada data em 2019.

	A	B
1	date	sales
2	1/1/2019	59.96
3	1/2/2019	67.06
4	1/3/2019	74.24
5	1/4/2019	78.11
6	1/5/2019	84.76
7	1/6/2019	100.60
8	1/7/2019	100.13
9	1/8/2019	96.36
10	1/9/2019	85.80

Você pode usá-los para determinar se uma data específica cai em um fim de semana ou feriado e adicionar uma coluna à sua planilha que reflita estas informações. Assim, você pode descobrir se as vendas nos fins de semana e feriados são maiores do que as vendas em outros dias. Isto será útil para saber para fins de planejamento de inventário e marketing.

PERGUNTA 4: COMO A RENTABILIDADE DIFERE PARA NOVOS CLIENTES EM RELAÇÃO AOS CLIENTES QUE RETORNAM?

Seu conjunto de dados não contém dados de vendas relacionados a novos clientes. Sem estes dados, você não será capaz de responder a pergunta final. Entretanto, pode acontecer que a empresa colete dados de clientes e os armazene em uma tabela de dados diferente.

Se assim for, seu próximo passo seria descobrir como acessar os dados dos clientes da empresa. Você pode então juntar os dados de receita de vendas à tabela de dados do cliente para categorizar cada venda como de um cliente novo ou retornando e analisar a diferença de lucratividade entre os dois conjuntos de clientes. Estas informações ajudarão suas partes interessadas a desenvolverem campanhas de marketing para tipos específicos de clientes a fim de aumentar a fidelidade à marca e a rentabilidade geral.

Conclusão

Ao trabalhar em projetos analíticos, você nem sempre terá à sua disposição todos os dados necessários ou relevantes. Em muitos desses casos, você pode recorrer a outras fontes de dados para preencher as lacunas.

Apesar das limitações de seu conjunto de dados, ainda é possível oferecer às suas partes interessadas algumas ideias valiosas. Para os próximos passos, seu melhor plano de ação será tomar a iniciativa de fazer perguntas, identificar outros conjuntos de dados relevantes, ou fazer alguma pesquisa por conta própria. Não importa com que dados você esteja trabalhando, inspecionar cuidadosamente seus dados causa um grande impacto na qualidade geral de sua análise.