

Formatar os dados iniciais

18 minutos

O Editor do Power Query no Power BI Desktop permite-lhe formatar (transformar) os seus dados importados. Pode efetuar ações como mudar o nome de colunas ou tabelas, alterar texto para números, remover linhas, definir a primeira linha como cabeçalho e muito mais. É importante formatar os seus dados para garantir que correspondem às suas necessidades e que são adequados para utilizar em relatórios.

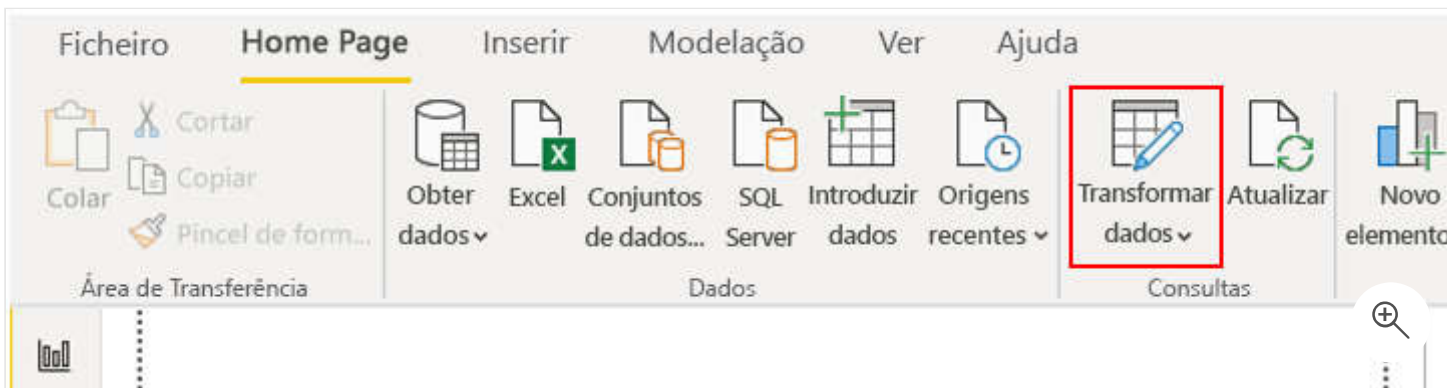
Carregou dados de vendas não processados de duas origens para um modelo do Power BI. Alguns dos dados foram importados de um ficheiro .csv criado manualmente no Microsoft Excel pela equipa de vendas. Os outros dados foram carregados através de uma ligação a um sistema de Planeamento de Recursos Empresariais (ERP) da sua organização. Agora, quando analisa os dados no Power BI Desktop, repara que estão desorganizados: alguns dados não são necessários e alguns dados necessários estão no formato errado.

Precisa de utilizar o Editor do Power Query para limpar e formatar estes dados antes de começar a criar relatórios.



Introdução ao Editor do Power Query

Para começar a formatar os seus dados, abra o Editor do Power Query ao selecionar a opção **Transformar dados** no separador **Base** do Power BI Desktop.



No Editor do Power Query, os dados na sua consulta selecionada são apresentados no meio do ecrã e, do lado esquerdo, o painel **Consultas** lista as consultas disponíveis (tabelas).

Quando trabalha no Editor do Power Query, todos os passos que segue para formatar os seus dados são registados. Depois, de cada vez que a consulta se liga à origem de dados, aplica automaticamente os seus passos, para que os dados estejam sempre formatados conforme especificou. O Editor do Power Query só faz alterações a uma vista particular dos seus dados, para que possa sentir-se confiante em relação às alterações que estão a ser aplicadas à sua origem de dados inicial. Pode ver uma lista dos seus passos no lado direito do ecrã, no painel **Definições da Consulta**, bem como as propriedades da consulta.

O friso do Editor do Power Query contém muitos botões que pode utilizar para selecionar, ver e formatar os seus dados.

Para saber mais sobre as funcionalidades e funções disponíveis, veja [The query ribbon](#) (O friso do Power Query).

📌 Nota

No Editor do Power Query, os menus de contexto e o separador **Transformar** no friso fornecem muitas das mesmas opções.

Identificar nomes e cabeçalhos de colunas

O primeiro passo na formatação dos seus dados iniciais é identificar os nomes e cabeçalhos das colunas nos dados e, em seguida, avaliar onde estão localizados para garantir que estão no local certo.

Na captura de ecrã seguinte, os dados de origem no ficheiro csv de SalesTarget (amostra não disponibilizada) tinham um destino categorizado por produtos e uma subcategoria dividida por meses, ambos organizados em colunas.

A	B	C	D	E	F	G	H	I	J	K	L	M	N
ProductSubcategoryID	Name	January	February	March	April	May	June	July	August	September	October	November	December
1	Mountain Bikes	780000	790000	800000	810000	820000	830000	840000	850000	860000	870000	880000	890000
2	Road Bikes	4500	5000	5500	6000	6500	7000	7500	8000	8500	9000	9500	10000
3	Touring Bikes	501000	502000	503000	504000	505000	506000	507000	508000	509000	510000	511000	512000
4	Handlebars	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200
5	Bottom Brackets	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200
6	Brakes	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200
7	Chains	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200
8	Cranksets	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200
9	Derailleurs	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000	2100	2200

No entanto, pode reparar que os dados não foram importados como esperado.

A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5	A ^B _C Column6	A ^B _C Column7
1		January	February	March	April	May
2						
3	ProductSubcategoryID	Name				
4	1	Mountain Bikes	780000	790000	800000	810000
5	2	Road Bikes	4500	5000	5500	6000
6	3	Touring Bikes	501000	502000	503000	504000
7	4	Handlebars	1100	1200	1300	1400
8	5	Bottom Brackets	1100	1200	1300	1400
9	6	Brakes	1100	1200	1300	1400
10	7	Chains	1100	1200	1300	1400
11	8	Cranksets	1100	1200	1300	1400
12	9	Derailleurs	1100	1200	1300	1400

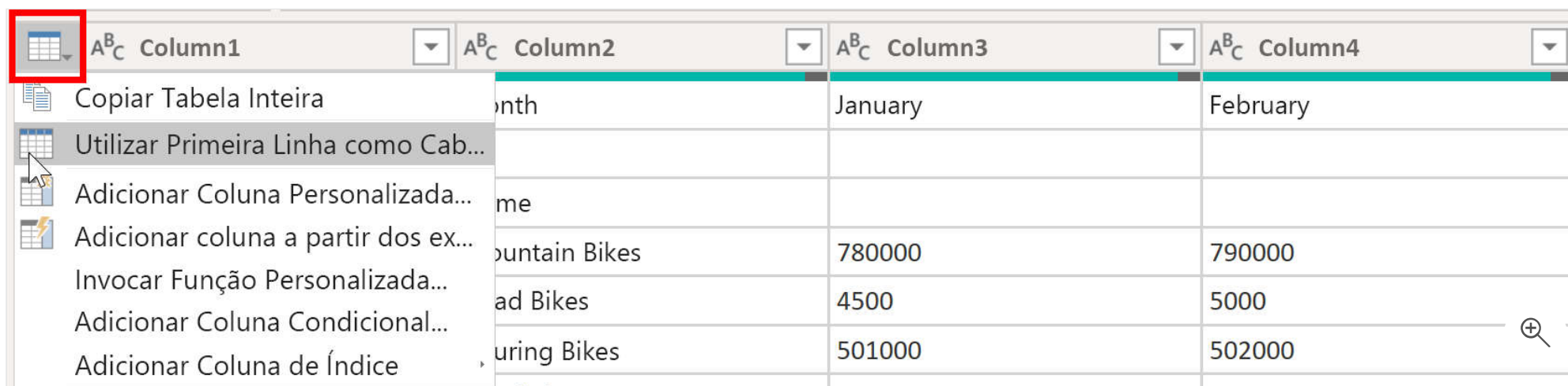
Consequentemente, os dados são difíceis de ler. Ocorreu um problema com os dados no seu estado atual, porque os cabeçalhos das colunas estão em linhas diferentes (marcadas a vermelho) e várias colunas têm nomes não descritivos, como **Column1**, **Column2**, entre outros.

Após identificar a localização dos nomes e cabeçalhos das colunas, pode fazer alterações para reorganizar os dados.

Promover cabeçalhos

Quando uma tabela é criada no Power BI Desktop, o Editor do Power Query assume que todos os dados pertencem a linhas da tabela. No entanto, uma origem de dados poderá ter uma primeira linha que contém nomes de colunas, o que aconteceu com o ficheiro SalesTarget do exemplo anterior. Para corrigir esta imprecisão, precisa de promover a primeira linha da tabela a cabeçalho da coluna.

Pode promover cabeçalhos de duas formas: ao seleccionar a opção **Utilizar Primeira Linha como Cabeçalhos** no separador **Base** ou ao seleccionar o botão de menu pendente junto a **Column1** e, em seguida, seleccionar **Utilizar Primeira Linha como Cabeçalhos**.




The screenshot shows the Power Query Editor interface. The top bar displays four columns: 'Column1', 'Column2', 'Column3', and 'Column4'. The 'Column1' dropdown menu is open, showing a list of actions. The option 'Utilizar Primeira Linha como Cabeçalhos' is highlighted. The data table below shows the following content:

Column1	Column2	Column3	Column4
January	February		
Mountain Bikes	780000	790000	
Ad Bikes	4500	5000	
Training Bikes	501000	502000	

A seguinte imagem ilustra a forma como a funcionalidade **Utilizar Primeira Linha como Cabeçalhos** afeta os dados:

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3	A ^B _C Column4	A ^B _C Column5
1		Month	January	February	March
2					
3	ProductSubcategoryID	Name			
4	1	Mountain Bikes	780000	790000	800000
5	2	Road Bikes	4500	5000	5500
6	3	Touring Bikes	501000	502000	503000
7	4	Handlebars	1100	1200	1300
8	5	Bottom Brackets	1100	1200	1300
9	6	Brakes	1100	1200	1300
10	7	Chains	1100	1200	1300



	A ^B _C	A ^B _C Month	A ^B _C January	A ^B _C February	A ^B _C March
1					
2	ProductSubcategoryID	Name			
3	1	Mountain Bikes	780000	790000	800000
4	2	Road Bikes	4500	5000	5500
5	3	Touring Bikes	501000	502000	503000
6	4	Handlebars	1100	1200	1300
7	5	Bottom Brackets	1100	1200	1300
8	6	Brakes	1100	1200	1300
9	7	Chains	1100	1200	1300
10	8	Cranksets	1100	1200	1300

Mudar o nome das colunas

O próximo passo na formação dos seus dados é examinar os cabeçalhos das colunas. Poderá descobrir que uma ou mais colunas têm os cabeçalhos errados, um cabeçalho tem um erro ortográfico ou a convenção de nomenclatura do cabeçalho não é consistente ou intuitiva.

Observe a captura de ecrã anterior, que mostra o impacto da funcionalidade **Utilizar Primeira Linha como Cabeçalhos**. Repare que a coluna que contém a subcategoria **Name** tem agora **Month** como cabeçalho. Este cabeçalho da coluna está incorreto, por isso é necessário mudá-lo o nome.

Pode mudar o nome dos cabeçalhos das colunas de duas formas. Pode clicar com o botão direito do rato no cabeçalho, seleccionar **Mudar o nome**, editar o nome e, em seguida, premir **Enter**. Em alternativa, pode clicar duas vezes no cabeçalho da coluna e substituir o nome com o nome correto.

Também pode resolver este problema ao remover (ignorar) as duas primeiras linhas e, em seguida, mudar o nome das colunas para o nome correto.

Remover as primeiras linhas

Ao formatar os seus dados, poderá precisar de remover algumas das primeiras linhas. Por exemplo, se estiverem em branco ou tiverem dados dos quais não precisa nos seus relatórios.

Se continuarmos com o exemplo do ficheiro SalesTarget, repare que a primeira linha está em branco (não tem dados) e a segunda linha tem dados que já não são necessários.

	ProductSubcategoryID	Subcategory Name	January	February	March
1					
2	ProductSubcategoryID	Name			
3	1	Mountain Bikes	780000	790000	800000
4	2	Road Bikes	4500	5000	5500
5	3	Touring Bikes	501000	502000	503000
6	4	Handlebars	1100	1200	1300

Para remover estas linhas em excesso, selecione **Remover Linhas** > **Remover Linhas Principais** no separador **Base**.

Home Page
Transformar
Adicionar Coluna
Ver
Ferramentas
Ajuda

Nova Origem
Origens Recentes
Introduzir Dados
Definições da origem de
Gerir Parâmetros
Atualizar Pré-Vis...

Propriedades
Editor Avançado
Gerir

Escolher Colunas
Remover Colunas
Manter Linhas
Remover Linhas

= Table.RenameColumns("#Promoted Headers",{{"Month", "Su

ProductSubcategoryID
Nome da Subcategoria
Janeiro

1
2
3

ProductSubcategoryID
Nome
Bicicletas de Montanha
780000

Remover Primeiras Linhas
Remover Últimas Linhas
Remover Linhas Alternadas
Remover Duplicados
Remover Linhas em Branco
Remover Erros

Remover colunas

Um passo fundamental no processo de formatação de dados é remover as colunas desnecessárias. O melhor é remover as colunas o mais cedo possível. Uma forma de remover as colunas seria ao limitar a coluna quando obtém os dados da origem de dados. Por exemplo, se estiver a extrair dados de uma base de dados relacional com o SQL, pode limitar a coluna que vai extrair ao utilizar uma lista de colunas na instrução SELECT.

Remover colunas numa fase inicial do processo é melhor do que numa fase mais avançada, especialmente quando estabeleceu relações entre as suas tabelas. Remover colunas desnecessárias irá ajudá-lo a concentrar-se nos dados necessários e a melhorar o desempenho geral dos seus conjuntos de dados e relatórios do Power BI Desktop.

Examine todas as colunas e verifique se precisa mesmo dos dados que contêm. Se não estiver a considerar utilizar esses dados num relatório, a coluna não acrescenta qualquer valor ao seu modelo de dados. Portanto, a coluna deve ser removida. Pode sempre adicionar a coluna mais tarde, se os seus requisitos se alterarem no futuro.

Pode remover colunas de duas formas. O primeiro método é ao seleccionar as colunas que pretende remover e, em seguida, no separador **Base**, seleccionar **Remover Colunas**.

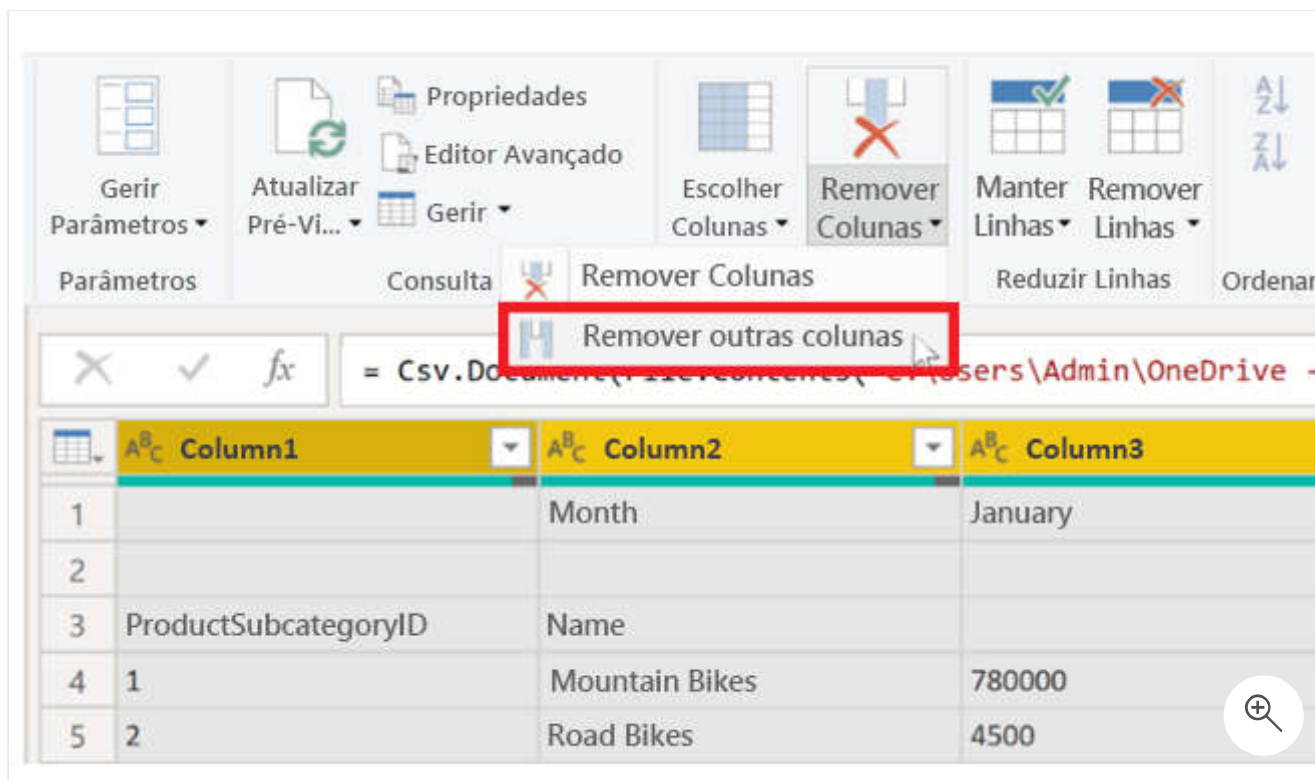
Propriedades
Editor Avançado
Manage ▾
Consulta

Escolher Colunas ▾
Remover Colunas ▾
Remover Colunas
Remover Outras Colunas

Manter Linhas ▾
Remover Linhas ▾
Reduzir Linhas
Ordenar

2	AB_C Coluna13	AB_C Coluna14
	Novembro	Dezembro
	880000	890000
	9500	10000
	511000	512000
	2100	2200
	2100	2200
	---	---

Em alternativa, pode seleccionar as colunas que pretende manter e, em seguida, no separador **Base**, seleccionar **Remover Colunas** > **Remover Outras Colunas**.



Anular dinamização das colunas

A opção Anular dinamização é uma funcionalidade útil do Power BI. Pode utilizar esta funcionalidade com dados de qualquer origem de dados, mas maioritariamente ao importar dados do Excel. O exemplo seguinte mostra um documento de exemplo do Excel com dados de vendas.

Year	2018	2019
January	\$ 15,370	\$ 16,063
February	\$ 15,950	\$ 12,161
March	\$ 13,862	\$ 14,180
April	\$ 18,530	\$ 6,516
May	\$ 5,203	\$ 19,395
June	\$ 5,928	\$ 19,324
July	\$ 14,736	\$ 15,939
August	\$ 6,243	\$ 15,390
September	\$ 15,178	\$ 17,832
October	\$ 18,148	\$ 5,185
November	\$ 8,014	\$ 9,222
December	\$ 19,470	\$ 14,000

Apesar de os dados fazerem sentido inicialmente, seria difícil criar um total de todas as vendas combinadas de 2018 e 2019. O seu objetivo seria utilizar estes dados no Power BI em três colunas: **Month** (Mês), **Year** (Ano) e **SalesAmount** (Montante de Vendas).

Ao importar os dados para o Power Query, ficará com um aspeto semelhante ao da imagem seguinte.

	A ⁸ C Year	1 ² ₃ 2018	1 ² ₃ 2019
1	January	15370	16063
2	February	15950	12161
3	March	13862	14180
4	April	18530	6516
5	May	5203	19395
6	June	5928	19324
7	July	14736	15939
8	August	6243	15390
9	September	15178	17832
10	October	18148	5185
11	November	8014	9290
12	December	19470	1

Em seguida, mude o nome da primeira coluna para **Month** (Mês). Esta coluna foi mal etiquetada porque o cabeçalho no Excel estava a etiquetar as colunas 2018 e 2019. Selecione as colunas 2018 e 2019, selecione o separador **Transformar** no Power Query e, em seguida, selecione **Anular dinamização**.

	^A _C Year	^A _C Attribute	¹ ₂ ₃ Value
	January	2018	15370
	January	2019	16063
	February	2018	15950
	February	2019	12161
	March	2018	13862
	March	2019	14180
	April	2018	18530
	April	2019	6516
	May	2018	5203
	May	2019	19395
1	June	2018	5928
2	June	2019	19324
3	July	2018	14736
4	July	2019	15939
5	August	2018	6243
5	August	2019	15390
7	September	2018	15178
8	September	2019	17832
9	October	2018	18148
0	October	2019	5185
1	November	2018	8014
2	November	2019	9299
3	December	2018	19470
4	December	2019	14087

Pode mudar o nome da coluna **Attribute** (Atributo) para **Year** (Ano) e a coluna **Value** (Valor) para **SalesAmount** (Montante de Vendas).

A opção Anular dinamização simplifica o processo de criação de medidas DAX nos dados posteriormente. Ao concluir este processo, criou uma forma mais simples de fragmentar os dados com as colunas **Year** (Ano) e **Month** (Mês).

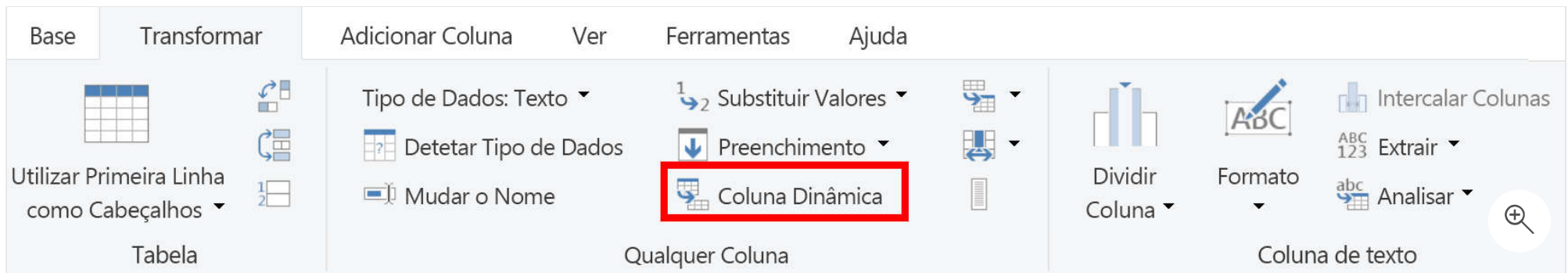
Colunas dinâmicas

Se os dados que está a formar são simples (ou seja, têm muitos detalhes, mas não estão organizados ou agrupados de forma específica), a falta de estrutura pode complicar a sua capacidade de identificar padrões nos dados.

Pode utilizar a funcionalidade **Coluna Dinâmica** para converter os seus dados simples numa tabela que contenha um valor agregado para cada valor exclusivo numa coluna. Por exemplo, poderá querer utilizar esta funcionalidade para resumir os dados ao utilizar diferentes funções matemáticas, como **Contagem**, **Mínimo**, **Máximo**, **Mediana**, **Média** ou **Soma**.

No exemplo de SalesTarget, pode dinamizar as colunas para obter a quantidade de subcategorias do produto em cada categoria de produto.

No separador **Transformar**, selecione **Transformar > Colunas Dinâmicas**.



Na janela **Coluna Dinâmica** apresentada, selecione uma coluna da lista **Coluna de Valores**, como **Nome da subcategoria**. Expanda as opções avançadas e selecione uma opção da lista **Função de Agregação de Valor**, como **Contagem (Tudo)** e, em seguida, selecione **OK**.

×

Coluna Dinâmica

Utilize os nomes na coluna "Category Name" para criar novas colunas.

Coluna de Valores ⓘ

Nome da Subcategoria ▾

▲ Opções avançadas

Função de Agregação de Valor ⓘ

Contagem (Tudo) ▾

[Saiba mais sobre a Coluna Dinâmica](#)

OK

Cancelar 🔍

A seguinte imagem ilustra a forma como a funcionalidade **Coluna Dinâmica** altera a forma como os dados são organizados.

	Category Name	Subcategory Name
1	Bikes	Mountain Bikes
2	Bikes	Road Bikes
3	Bikes	Touring Bikes
4	Clothing	Bib-Shorts
5	Clothing	Caps
6	Clothing	Gloves
7	Clothing	Jerseys
8	Clothing	Shorts
9	Clothing	Socks
10	Clothing	Tights
11	Clothing	Vests
12	Accessories	Bike Racks
13	Accessories	Bike Stands
14	Accessories	Bottles and Cages

	1.2 Bikes	1.2 Components	1.2 Clothing	1.2 Accessories
1	3	14	8	12

O Editor do Power Query regista todos os passos seguidos para formatar os dados e a lista de passos é apresentada no painel **Definições da Consulta**. Se tiver feito todas as alterações necessárias, selecione **Fechar e Aplicar** para fechar o Editor do Power Query e aplicar as suas alterações ao modelo de dados. No entanto, antes de seleccionar **Fechar e Aplicar**, pode seguir mais passos para limpar e transformar os seus dados no Editor do Power Query. Estes passos adicionais serão abordados mais à frente neste módulo.

Unidade seguinte: Simplificar a estrutura de dados

Continuar >