

## Exercício 1 12 – Remover linhas não desejadas<sup>1</sup>

Quando são importados dados para o “Power BI” é comum não começarem a ser apresentados na primeira linha. Se os dados foram gerados automaticamente por outro sistema, podem apresentar informação que não seja necessária para a sua análise (por exemplo, títulos ou metadados).

a) Seguindo o caminho “home → get data” carregue os dados armazenados no ficheiro “exercicio\_1\_12\_a.csv”. Vamos editar os dados no “Power Query Editor” e verificar que temos o título dos dados na linha 1 e, em seguida, uma linha em branco antes do início dos dados. Podemos remover estas duas linhas usando o comando “home → remove rows → remove top rows”. De seguida devemos promover os cabeçalhos, seguindo “home → use first row as header”. Concluído o exercício, grave o ficheiro criado no “Power BI” com a designação “exercicio\_1\_12\_a (resolvido).pbix”.

b) Uma outra situação frequente, em que é necessário remover linhas, acontece quando se importa uma tabela do Excel que utiliza o totalizador de linhas. Pode verificar-se que na parte inferior da tabela “exercicio\_1\_12\_b.xlsx” há uma linha total (esta linha pode ser ativada e desativada, no Excel, em “table design”). Assim, sempre que se importa uma tabela do Excel, vale a pena verificar se há alguma linha de total ativa.

Seguindo o caminho “home → get data” vamos importar o ficheiro “exercicio\_1\_12\_b.xlsx”. Se tivermos a certeza de que a linha total está sempre ativa, podemos usar o comando “remove rows”. No entanto, podemos usar um método que oferece maior flexibilidade para remover linhas de dados indesejadas. Verificando se na primeira coluna existe a palavra “total”, podemos simplesmente clicar na caixa de seleção (filtros) e desativá-la. Com esta abordagem não corremos o risco de excluir uma linha de dados que possa vir a ser necessária numa outra altura. De forma semelhante, poderíamos em qualquer uma das colunas com valores numéricos usar filtros (por exemplo, maior, maior ou igual a, ou entre). Concluído o exercício, grave o ficheiro criado no “Power BI” com a designação “exercicio\_1\_12\_b\_R.pbix”.

Date	Product	Serial number	Sale price	Cost price	Telesales time (mins)	Preparation time (mins)
27/03/2017	DX103	GF54542	68.93	35.16	33	14
19/03/2017	DX103	GF54286	68.93	35.84	3	21
17/03/2017	AX102	GF54268	51.28	26.15	17	14
24/03/2017	AX101	GF54553	68.71	32.30	40	23
27/03/2017	AX103	GF54373	53.50	28.36	27	11
03/03/2017	DX103	GF54602	68.93	35.16	17	9
23/03/2017	DX101	GF54599	17.62	8.64	33	16
10/03/2017	DX103	GF54456	68.93	37.22	0	18
27/03/2017	AX102	GF54322	51.28	24.62	28	17
29/03/2017	DX102	GF54356	62.46	32.48	30	13
21/03/2017	AX102	GF54301	51.28	25.64	15	9

Date	Product	Serial number	Sale price	Cost price	Telesales time (mins)	Preparation time (mins)
06/03/2017	DX101	GF54552	17.62	9.34	8	22
06/03/2017	DX102	GF54679	24.79	11.40	25	9
06/03/2017	DX102	GF54251	53.57	25.18	19	21
27/03/2017	DX103	GF54642	68.93	35.16	33	14
19/03/2017	DX103	GF54286	68.93	35.84	3	21
17/03/2017	AX102	GF54268	51.28	26.15	17	14
24/03/2017	AX101	GF54553	68.71	32.30	40	23
27/03/2017	AX103	GF54373	53.50	28.36	27	11
03/03/2017	DX103	GF54602	68.93	35.16	17	9
23/03/2017	DX101	GF54599	17.62	8.64	33	16
10/03/2017	DX103	GF54456	68.93	37.22	0	18
27/03/2017	AX102	GF54322	51.28	24.62	28	17
29/03/2017	DX102	GF54356	62.46	32.48	30	13
21/03/2017	AX102	GF54301	51.28	25.64	15	9
Total			5,352.20	2,519.62		

<sup>1</sup> Gamble, G. (2018). *Power BI Step-by-Step Part 3: Power Query, Parameters, Templates & Custom Functions*

## Exercício 1\_13 – Substituir valores e preencher<sup>2</sup>

Neste exercício, vamos tomar contacto com os comandos “**replace value**” e “**fill down**”. Estes comandos são muito úteis quando se pretende **modificar dados**, relativamente ao conteúdo original. Para exemplificar a sua utilização vamos importar o ficheiro “**exercicio\_1\_13.txt**”. Clique em “**home → get data → text/CSV**” e abra os dados no “**Power Query Editor**”. Devemos começar por **remover duas linhas indesejadas e usar a primeira linha como cabeçalhos**.

Para converter a coluna “**amount**” no formato moeda, vamos **retirar o símbolo monetário da libra**, **alterar o carácter decimal para vírgula** e utilizar o “**fixed decimal number**”. Repare que permanecem **dois erros**. Isto acontece, porque existem duas transações que ainda não foram validadas. Podemos excluir estas transações ou substituir o erro por um valor numérico. Neste caso, vamos optar por **excluir as transações**. ?

Verifica-se que todas as categorias estão escritas por extenso com exceção da categoria “**Miscellaneous**”, que está representada abreviadamente. Para manter a coerência, vamos substituir “**Misc.**” por “**Miscellaneous**” (“**replace value**”). ✓

Finalmente, vamos preencher os campos vazios pela designação da categoria. Isto pode ser feito recorrendo ao comando “**fill down**”. Este comando copia cada um dos títulos (na coluna categoria) e repete-os nas linhas em branco, imediatamente abaixo, nunca sobrescrevendo uma entrada (somente os campos vazios serão preenchidos).

No entanto, o comando “**fill down**” apenas substitui valores nulos e o “**Power BI**” não considera uma entrada em branco como nula. Portanto, se tentarmos preencher as células em branco, clicando com o botão direito do rato na coluna “categoria” e escolhendo “**fill down**” no menu de contexto, veremos que, embora a etapa “**fill down**” seja criada, nada acontece com os dados, os **espaços em branco mantêm-se**. Isto acontece, **porque o conteúdo das células que seguem cada entrada não são valores nulos**, eles estão em branco. Assim, antes de utilizar o comando “**fill down**”, devemos **substituir essas células em branco por nulos (null)** (“**replace value**”). ✓  
Depois de substituir os brancos por nulos, podemos usar de novo o comando “**fill down**” e obter o resultado desejado.

Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_13\_R.pbix**”.

Date Claimed	Amount
03/10/2016	103,01
07/10/2016	29,16
11/01/2016	1,02
26/12/2016	62,04
09/12/2016	526,99
11/11/2016	366,45
11/03/2016	461,91
17/06/2016	Not agreed
20/12/2016	363,57
14/01/2016	67,41
08/03/2016	269,31
27/05/2016	150,15
12/02/2016	376,77
08/01/2016	383,91
22/01/2016	508,74
11/04/2016	387,47
06/06/2016	15,48

### Exercício 1 14 – Anular a dinamização de colunas<sup>3</sup>

Há alturas em que os dados que pretendemos analisar estão devidamente estruturados e “limpos”. No entanto, não estão no formato normalizado que pretendemos. É o que acontece com o ficheiro “exercício\_1\_14.xlsx”.

Importe o ficheiro e verifique no “Power Query Editor” que temos uma lista de restaurantes com as classificações atribuídas pelos seus clientes. As classificações são organizadas em categorias. No entanto, essas categorias foram colocadas em colunas separadas. Esta colocação dificulta a criação de um relatório, com base nestes dados, designadamente se pretendermos fazer uma análise comparativa da popularidade dos diferentes restaurantes. Nestas situações, a remoção da dinamização das colunas torna-se útil. Há duas maneiras de proceder: selecionar as colunas que se deseja remover da dinamização ou, se estas forem a minoria, pode ser mais rápido selecionar as colunas restantes.

Os atributos “restaurante” e “data” correspondem às duas únicas colunas em que não pretendemos anular a dinamização. Todas as outras colunas se relacionam com a avaliação, devendo ser anulada a sua dinamização. Podemos encontrar os comandos “unpivot columns” e “unpivot other columns” clicando com o botão direito do rato nas colunas selecionadas e selecionar a opção no menu de contexto ou indo para a o separador “transform”.

seleciono colunas "restaurant" e "date" e faço "unpivot other columns"  
Pois estas duas n se pretende anular a dinamização

Quando se anula a dinamização de colunas, obtêm-se duas colunas, chamadas “atributo” e “valor”. A coluna “atributo” contém os títulos das colunas originais, enquanto a coluna “valores” contém as entradas que estavam abaixo de cada um desses títulos. O “Power Query Editor” combina tudo automaticamente, criando as necessárias linhas extra.

Finalmente, devemos renomear essas duas colunas resultantes. Então, vamos substituir “Atributo” por “Category” e “Valor” por “Classification”. ✓

Agora temos um conjunto de dados mais flexível que facilitará a construção de relatórios. Realizando uma verificação final, podemos verificar que temos um número inteiro para a classificação e uma data para a data, podemos então clicar em “fechar e aplicar”.

Concluído o exercício, grave o ficheiro criado no “Power BI” com a designação “exercicio\_1\_14\_R.pbix”. ✓

<sup>3</sup> Gamble, G. (2018). *Power BI Step-by-Step Part 3: Power Query, Parameters, Templates & Custom Functions*

### Exercício 1\_15 – Reordenar colunas<sup>4</sup>

Para este exercício, vamos **importar os dados armazenados** no ficheiro “**exercicio\_1\_15.xlsx**”. Este ficheiro, como se pode verificar, contém **uma única folha** de trabalho do Excel, e **não uma tabela**. Nestas situações pode acontecer serem importadas colunas extras em branco como parte do conjunto de dados. Portanto, além de **reordenar as colunas**, podemos também ter necessidade de **remover algumas colunas**.

Como já nos apercebemos nos exercícios realizados anteriormente, o “**Power Query Editor**” introduz algumas etapas de **forma automática** para nos ajudarem, no entanto, esta ajuda pode ser um pouco prematura, pelo que podemos descartar esta etapa. Se o pretender, pode impedir que o “**Power Query Editor**” insira automaticamente estas etapas de ajuda seguindo o caminho “**file → options and settings → options**” e em seguida, na seção “**current file**”, clique em “**data load**” e desative a opção “**detect column types and headers for unstructured sources**”.

Após importar o ficheiro “**exercicio\_1\_15.xlsx**” e no “**Power Query Editor**” pode verificar que da coluna 19 à coluna 23 não tem conteúdo, pelo que as pode eliminar. Pressione a tecla “**Shift**” e clique nos cabeçalhos da **primeira e última coluna que pretende eliminar**. Seguidamente, clique com o botão direito do rato no topo de uma das colunas seleccionadas e no menu de contexto selecione a opção “**remove columns**” ou, de maneira alternativa, pode encontrar a mesma funcionalidade no separador “**home**”.

Existem duas técnicas principais para reordenar colunas. A primeira técnica é mover uma ou mais colunas seleccionadas para uma posição específica. Mover, por exemplo, coluna “**bloodtype**” para aparecer após as colunas “**weight**” e “**height**”. A segunda técnica consiste em mover uma ou mais colunas em relação às restantes colunas. Por exemplo, **se quisermos tratar o “emailaddress” como a coluna-chave**, podemos colocá-la como primeira coluna. Para isso basta seleccionar a coluna e seguidamente, clicando com o botão direito, seleccionar “**move**” e escolher uma das opções disponíveis. Pode, em alternativa, optar por seguir o separador “**transform**”.

Também podemos pretender que as duas colunas relativas ao cartão de crédito “**CCType**” e “**CCExpire**” apareçam no início, imediatamente após a coluna “**emailaddress**”. Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_15\_R.pbix**”.

<sup>4</sup> Gamble, G. (2018). *Power BI Step-by-Step Part 3: Power Query, Parameters, Templates & Custom Functions*

## Exercício 1 16 – Criar colunas personalizadas<sup>5</sup>

Vamos retomar o exercício anterior para criarmos colunas personalizadas. Estamos interessados em criar uma **coluna que apresente a altura, em centímetros**, e uma **coluna que apresente o índice de massa corporal (BMI)**. O **BMI pode ser calculado a partir do peso, em libras, e da altura, em polegadas**. Precisamos, então, de criar uma nova coluna relativa à altura, atualmente a coluna **“height”** é uma coluna de texto que contém pés e polegadas (por exemplo, 5'11"). Temos necessidade de criar uma nova coluna relativa à altura, **contendo apenas polegadas**, que pode então ser convertida num tipo de dados numérico. Vamos dividir a coluna **“height”**, usando o **apóstrofo como delimitador**. Seleccionamos a coluna, clicamos com o botão direito e seleccionamos **“split column”**.

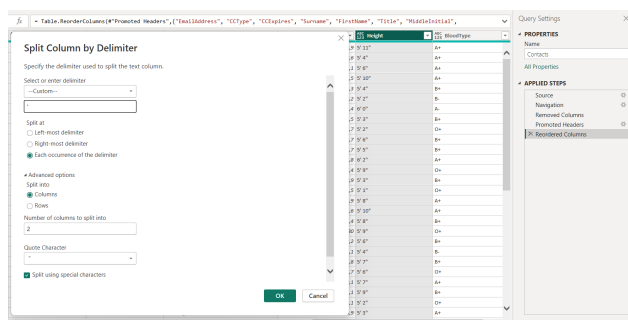
Como se pode verificar, criámos duas colunas, sendo agora necessário criar uma única coluna por junção destas duas. Assim, clicamos no tab **“add column → custom column”**. Vamos designar a nova coluna **“Height”**. Em seguida, clicamos na caixa da fórmula e **seleccionamos as colunas adequadas para personalizar a nova coluna**. De seguida pode remover as colunas **“Height.1”** e **“Height.2”**.

Agora podemos criar o BMI, clicando em **“add column → custom column”** e introduzir a designação **“BMI”**.

O BMI, no sistema inglês, pode ser obtido através da seguinte expressão:

$$\text{BMI} = (\text{weight in pounds} / (\text{height in inches} * \text{height in inches})) * 703$$

Concluído o exercício, grave o ficheiro criado no **“Power BI”** com a designação **“exercicio\_1\_16\_R.pbix”**.



### Custom Column

Add a column that is computed from the other columns.

New column name

Height

Custom column formula

= [Height.1] \* 30.48 + [Height.2]\*2.54

Available columns

EmailAddress  
CCType  
CCExpires  
Surname  
FirstName  
Title  
MiddleInitial

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

### Custom Column

Add a column that is computed from the other columns.

New column name

BMI

Custom column formula

= ([Weight] / ([Height] / 2.54 \* [Height] / 2.54)) \* 703

Available columns

EmailAddress  
CCType  
CCExpires  
Surname  
FirstName  
Title  
MiddleInitial

<< Insert

[Learn about Power Query formulas](#)

✓ No syntax errors have been detected.

OK

Cancel

<sup>5</sup> Gamble, G. (2018). *Power BI Step-by-Step Part 3: Power Query, Paramete*