

## Exercício 1\_07 - Importar dados de uma pasta<sup>1</sup>

Um cenário bastante comum, nas organizações, consiste em importar dados com uma determinada periodicidade. Admita, por exemplo, que é o responsável pelo departamento de vendas da empresa “**Vende Bem, SA**” e que pretende fazer uma análise anual da evolução das suas vendas. Por norma, colocará numa pasta o mapa mensal de vendas, pretendendo que o ficheiro, anual, no “**Power BI**” seja automaticamente atualizado. O ficheiro inicial “**exercicio\_1\_07\_jan.csv**”, relativo ao mês de janeiro, é trabalhado na pasta “**exercicio\_1\_07 (mapas\_em\_curso)**”. Quando fechado, este ficheiro, é transferido para a pasta “**exercicio\_1\_07 (mapas\_fechados)**”, dando-se início ao tratamento de um novo ficheiro relativo às vendas realizadas no mês de fevereiro “**exercicio\_1\_07 fev.csv**”.

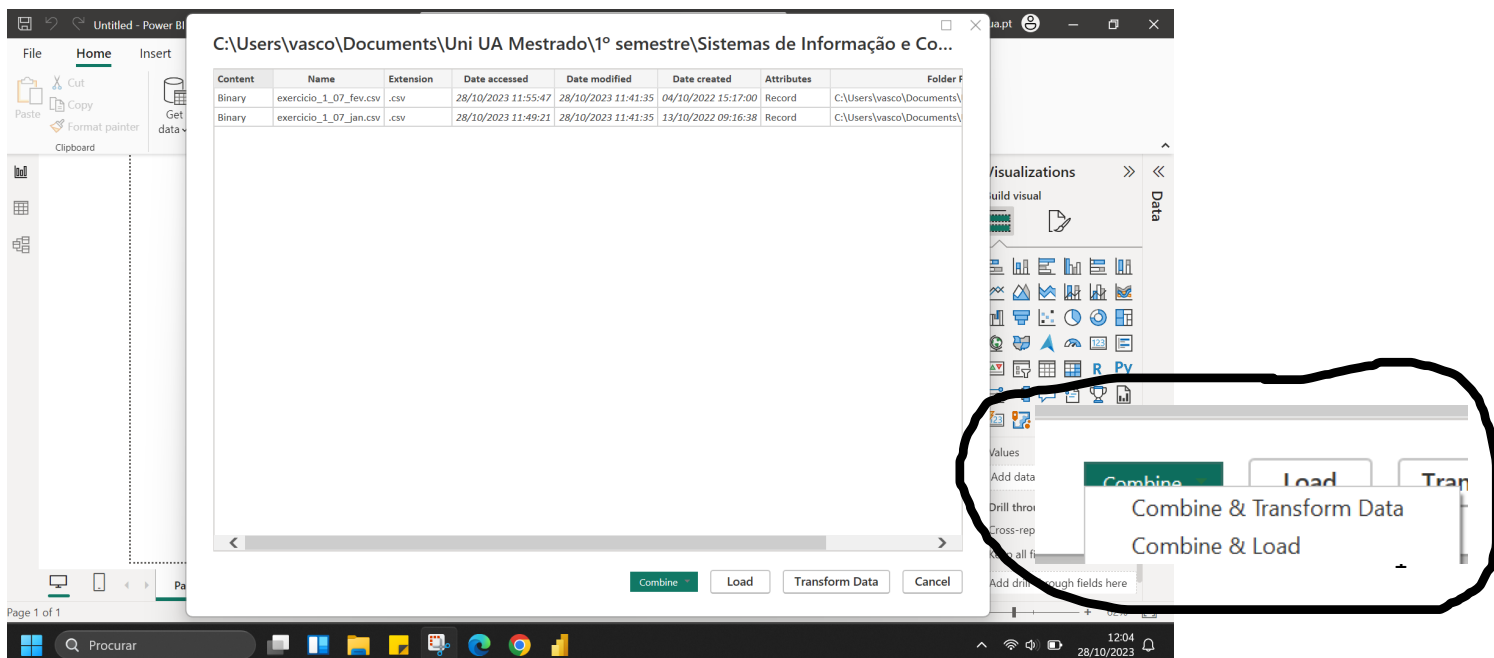
Vamos carregar no “**Power BI**” o ficheiro “**exercicio\_1\_07\_jan.csv**”, depois de colocado na pasta “**exercicio\_1\_07 (mapas\_fechados)**”, seguindo o caminho “**home → get data → folder**”, e proceder a algumas transformações nos dados importados: eliminar a primeira coluna; reorganizar colunas (colocar a data no início) e converter a margem para valores decimais.

Depois de importado e tratado o ficheiro, relativo ao mês de janeiro, vamos atualizar a pasta “**exercicio\_1\_07 (mapas\_fechados)**” com cada um dos restantes ficheiros mensais, provenientes da pasta “**exercicio\_1\_07 (mapas\_em\_curso)**”, à medida que vão sendo fechados. Não esquecer de ir fazendo o refresh do “**Power BI**” por cada ficheiro transferido para a pasta “**exercicio\_1\_07 (mapas\_fechados)**”.

Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_07\_R.pbix**”.

Criar pasta (mapas fechados), importar dados de pasta, editar o ficheiro em "query edit", às vezes a pasta não muda de lugar. Carregar em "replace value" e trocar "ponto" por "virgula", mudar o "data type" para "decimal number".

Depois de ter feito a primeira ligação é só colocar ficheiros na base de dados queo power bi vai atualizando



### Exercício 1\_09 – Editar dados e usar as funções “trim” e “clean”<sup>3</sup>

Continuando a analisar, situações com que os utilizadores de dados organizacionais se possam deparar com frequência, vamos considerar a situação em que é necessário remover espaços que possam existir em vários campos de dados, após a sua importação para o “Power BI” e que possam ter impacto na sua análise.

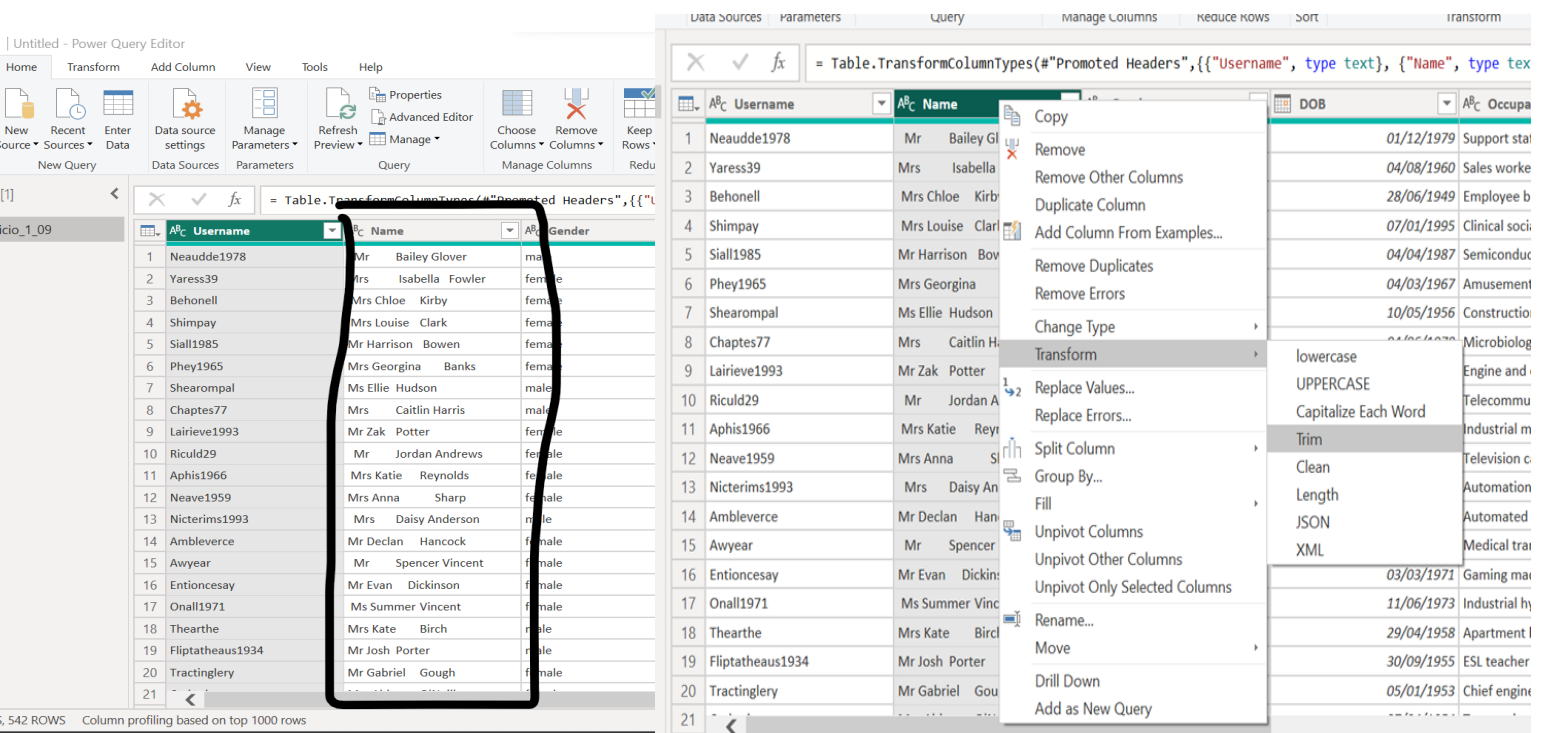
Vamos importar para o “Power BI” o ficheiro “exercicio\_1\_09.csv”. Para isso podemos seguir o seguinte caminho: “home → get data → Text/CSV” e seleccionar o ficheiro desejado. Como pretendemos fazer alterações aos dados importados, vamos trabalhar com o “Power Query Editor”, para isso, clicamos no botão “transform data”.

O “Power Query Editor”, oferece um ambiente muito rico e poderoso com algumas funções de limpeza muito úteis. Olhando para os dados na coluna “name” podemos verificar que os mesmos apresentam problemas relacionados com a existência de espaços extra: espaços precedendo e sucedendo algumas entradas e também entre palavras.

Para corrigir estas anomalias podemos recorrer à função “trim”, que remove todos os espaços em branco à esquerda e à direita do texto “transform → format → trim”, e à função “clean”, que devolve o texto com todos os caracteres não imprimíveis removidos “transform → format → clean”. Após estas transformações verificamos que ainda subsistem espaços em branco que podem ser resolvidos recorrendo ao comando “home → replace values”.

Vamos ainda proceder a uma alteração que obrigue a primeira letra da coluna “gender” a ser maiúscula “transform → format → capitalize each word”. Após concluir estas transformações, grave o ficheiro com a designação “exercicio\_1\_09\_R.pbix”.

Fazer 1 trim, 1 clean e fazer 4 replaced por vazio



The screenshot displays the Power Query Editor interface. The top ribbon includes tabs for Home, Transform, Add Column, View, Tools, and Help. The main area shows a data table with columns: Username, Name, Gender, and DOB. The 'Name' column contains names with extra spaces, such as 'Mr Bailey Glover' and 'Mrs Isabella Fowler'. A context menu is open over the 'Name' column, showing options like 'Copy', 'Remove', 'Duplicate Column', 'Add Column From Examples...', 'Remove Duplicates', 'Remove Errors', 'Change Type', 'Transform', 'Replace Values...', 'Replace Errors...', 'Split Column', 'Group By...', 'Fill', 'Unpivot Columns', 'Unpivot Other Columns', 'Unpivot Only Selected Columns', 'Rename...', 'Move', 'Drill Down', and 'Add as New Query'. The 'Transform' option is highlighted, and a sub-menu is visible showing 'lowercase', 'UPPERCASE', 'Capitalize Each Word', 'Trim', 'Clean', 'Length', 'JSON', and 'XML'. The 'Trim' option is selected. The bottom status bar indicates '542 ROWS' and 'Column profiling based on top 1000 rows'.



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

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

New Source ▾	Recent Sources ▾	Enter Data	Data source settings	Manage Parameters ▾	Refresh Preview ▾	Manage ▾	Choose Columns ▾	Remove Columns ▾
New Query			Data Sources	Parameters		Query	Manage Columns	

fx	= Table.TransformColumnTypes(Responses_Table,{{"I	
----	---	--

ABc Restaurant	Date	Overall
1 Spicy	05/07/2016	
2 Spring	15/04/2016	
3 Ember	26/07/2016	
4 Ember	10/11/2016	
5 Ember	22/03/2016	
6 Spicy	21/05/2016	
7 Spicy	10/07/2016	
8 Ember	26/07/2016	
9 Benga	03/01/2016	
10 Ember	15/05/2016	
11 Benga	27/08/2016	
12 Spicy	31/05/2016	
13 Spicy	20/03/2016	
14 Benga	23/11/2016	
15 Benga	01/10/2016	
16 Bengal Petal	04/11/2016	
17 Mellow Spices	29/10/2016	
18 Spicy View	15/03/2016	
19 Embers Acton	12/11/2016	
20 Embers Ealing	24/02/2016	

Clicar em "control" + "botão esquerdo do rato" para seleccionar as colunas base, depois de seleccionadas clicar no botão direito e seleccionar unpivot outhter colluns e o power bi vai reorganizar as outras colunas para que apareça a avaliação a cada categoria por restaurante

<sup>3</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**”.

The screenshot shows the Power BI Desktop interface. On the left, a data table is visible with columns 'Weight' and 'Height'. The 'Height' column contains values like '198,9 5' 11"', '167,6 5' 4"', etc. A right-click context menu is open over the 'Height' column, with 'Split Column' selected. On the right, the 'Custom Column' dialog box is open. It shows the 'New column name' as 'Custom' and the 'Custom column formula' as `= [Weight] / ((([Height.2] + [Ftoots to Inches])) * ([Height.2] + [Ftoots to Inches])) * 703`. The 'Available columns' list on the right includes 'EmailAddress', 'CCType', 'CCExpires', 'Surname', 'FirstName', 'Title', 'MiddleInitial', and 'Country'. The 'OK' button is highlighted.



### Exercício 1 17 – Consolidar consultas (acrescentar)<sup>1</sup>

No “**Power BI**” existe um recurso, bastante útil, que permite juntar linhas com dados correspondentes: “**Append Queries**”.

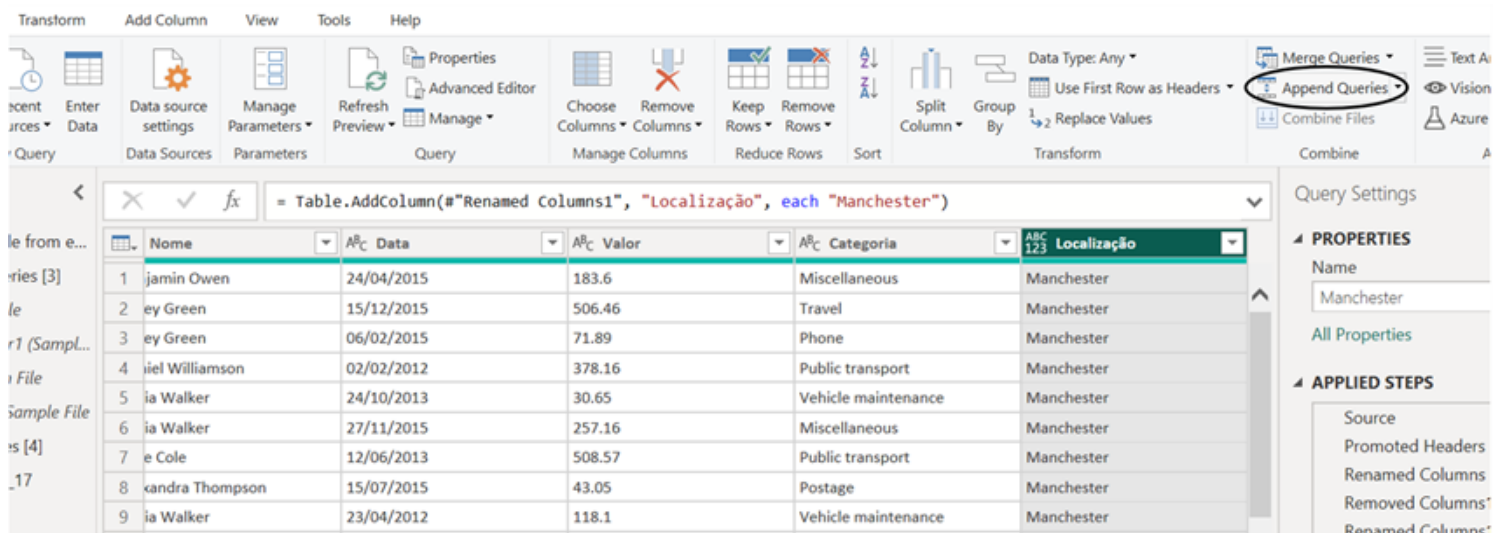
Começamos por importar os dados armazenados nos ficheiros CSV guardados na pasta “**exercicio\_1\_17**”. Faça um duplo clique no ficheiro “**Leeds.csv**”, carregando os seus dados no “**Editor do Power Query**”. Seguidamente, dentro do “**Editor do Power Query**”, clique em “**Home → New Source → Text/CSV**” e clique duas vezes em “**Liverpool.csv**” para importar o ficheiro. Repita o processo para carregar “**Manchester.csv**”.

Os três ficheiros, acabados de importar, contêm, no essencial, o mesmo tipo de dados. Devemos começar por remover as diferenças entre as colunas, fazendo com que tenham os mesmos cabeçalhos e o mesmo tipo de dados nas três consultas. Começando o processo com a consulta “**Leeds**”, vamos atribuir as seguintes designações às colunas: “**Nome**”; “**Data**”; “**Valor**” e “**Categoria**”. Seguidamente vamos alterar as consultas “**Liverpool.csv**” e “**Manchester.csv**”, em conformidade. Na consulta “**Liverpool**”, somente, será necessário renomear as colunas. Na consulta “**Manchester**” será necessário renomear as colunas e remover a coluna “**Date Paid**” (não existe nas outras consultas).

Poderá ser útil acrescentar uma coluna que identifique o local em que cada pessoa trabalha. Portanto, adicionaremos uma coluna “**Localização**” a cada uma das três consultas. Faremos isto, seguindo o caminho “**Add column → custom column**” e atribuindo a designação “**Localização**” à nova coluna com o texto “**Leeds**”, “**Liverpool**” ou “**Manchester**”, conforme o caso. Tratadas as consultas, podemos acrescentar uma nova consulta que faça a sua consolidação, seguindo o seguinte caminho “**Home → (combine) → append queries → append queries as new**”.

Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_17\_R.pbix**”.

Ter atenção ao título das colunas, tem de ser igual em todas as folhas,  
Tem de haver uma coluna que ligue as 3 folhas



	Nome	Data	Valor	Categoria	Localização
1	Jamin Owen	24/04/2015	183.6	Miscellaneous	Manchester
2	Ey Green	15/12/2015	506.46	Travel	Manchester
3	Ey Green	06/02/2015	71.89	Phone	Manchester
4	Niel Williamson	02/02/2012	378.16	Public transport	Manchester
5	ia Walker	24/10/2013	30.65	Vehicle maintenance	Manchester
6	ia Walker	27/11/2015	257.16	Miscellaneous	Manchester
7	e Cole	12/06/2013	508.57	Public transport	Manchester
8	kandra Thompson	15/07/2015	43.05	Postage	Manchester
9	ia Walker	23/04/2012	118.1	Vehicle maintenance	Manchester

## Exercício 1 18 – Consolidar consultas (intercalar)<sup>2</sup>

No exercício anterior, vimos como o recurso “**append queries**” pode ser usado para juntar linhas de dados de diferentes consultas. Neste exercício, veremos um outro recurso similar, “**merge queries**”, que é usado para juntar colunas de dados de diferentes consultas.

Vamos considerar que temos duas fontes de dados, uma contém informação pública e outra contém informação confidencial. Neste caso, temos dados sobre utentes de uma unidade de saúde: no ficheiro “**Data.xlsx**” temos dados públicos e no ficheiro “**Personal.xlsx**” temos dados privados. Pretendemos reunir estes dois ficheiros que contém dados sobre os mesmos indivíduos, identificados pela coluna de endereço de e-mail. A utilização do recurso “**merge queries**” carece de uma coluna comum a ambas as consultas. Após tratar os dados importados e renomear as consultas de acordo com o tipo de dados, podemos proceder à sua intercalação, seguindo o seguinte caminho “**Home → (combine) → merge queries → merge queries as new**”.

Selecionada a consulta “Dados Públicos”, na primeira caixa de diálogo será apresentada esta consulta, sendo agora necessário especificar que o endereço de e-mail é a coluna com a qual queremos fazer a correspondência (coluna comum a ambas as consultas). Isto é feito clicando simplesmente na coluna respetiva. De seguida, indicamos a segunda consulta, “Dados Privados”, com a qual pretendemos fazer a intercalação e indicamos a coluna com a qual queremos fazer a correspondência (endereço de e-mail). Neste caso, pretendemos apenas que as colunas correspondentes de ambas as consultas sejam intercaladas. Assim, escolhemos a opção **inner** (*only matching rows*). Finalmente, clicamos em OK e é criada uma nova consulta que contém os dados públicos e uma coluna com os dados privados que é necessário expandir, desativando a opção “*use original column name as prefix*”, que neste caso não tem utilidade. Também podemos desativar a coluna do endereço de e-mail ou então eliminá-la à posteriori, depois de confirmar que a intercalação foi bem realizada.

Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_18\_R.pbix**”.

**Home**

**Merge Queries**

**Merge Queries as New**

**Combine**

**Query Settings**

**PROPERTIES**

**Name**

**Merge**

Select tables and matching columns to create a merged table.

EmailAddress (Personal)

EmailAddress	Occupation	BloodType	Weight	Height	CCType	CCEXpires
BaileyGlover@gustr.com	Support staff clerk	A+	198,9	5' 11"	MasterCard	31/12/2017
IsabellaFowler@cuvov.de	Financial consultant	A+	167,6	5' 4"	Visa	30/06/2017
ChloeKirby@gustr.com	Sales worker supervisor	A+	184,1	5' 6"	Visa	30/04/2018
MohammadGilbert@cuvov.de	Pharmacist	A+	205,5	5' 10"	MasterCard	31/01/2014

EmailAddress (General Data)

EmailAddress	Surname	FirstName	Title	MiddleInitial	StreetAddress	City
BaileyGlover@gustr.com	Glover	Bailey	Mr.	K	70 Whitchurch Road	Emery Down
IsabellaFowler@cuvov.de	Fowler	Isabella	Mrs.	C	41 Rowland Rd	Otterburn
ChloeKirby@gustr.com	Kirby	Chloe	Mrs.	F	60 Graham Road	Chenies
MohammadGilbert@cuvov.de	Gilbert	Mohammad	Mr.	L	87 Coast Rd	Kirknewton

Join Kind

Inner (only matching rows)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

✓ The selection matches 1106 of 1106 rows from the first table, and 1106 of...

**OK** **Cancel**

<sup>2</sup> Gamble, G. (2018). *Power BI Step by Step*. Microsoft Press.

Dora Simões & Carlos Seabra

### Exercício 1 19 – Agrupar dados<sup>3</sup>

Quando utilizamos fontes de dados que possuem muitas linhas, e mais linhas do que são realmente necessárias, pode ser útil agregar os dados antes de proceder à sua importação. O comando *Group by* do Query Editor permite-nos fazer isso.

Vejamos como podemos utilizar o recurso indicado anteriormente, quando ligamos a uma pasta de trabalho do Excel. Seguindo o caminho “**Home → get data → Excel workbook**” navegaremos até à pasta “**exercicio\_1\_19**” e façamos um duplo clique em “**All Expenses Data.xlsx**” e entremos no “**Power Query Editor**”. Os dados guardados neste ficheiro são relativos ao detalhe de cada reclamação feita aos diversos gestores de conta. O nível de granularidade apresentado pode não ser necessário, sendo suficiente conhecer o total reclamado a cada gestor de conta. Assim, os dados podem ser agrupados pela coluna “**Name**” em que é mostrado o valor reivindicado junto de cada um dos gestores de conta.

Para usar o recurso de agrupar dados, basta destacar a coluna “**Name**” e clicar em “**Home → group by**”. Como seleccionamos a coluna “**Name**”, o nome é mostrado como grupo por coluna. Seguidamente, vamos especificar o nome da nova coluna, pode ser “**Valor**”, e a operação que queremos é somar valores da coluna “**Amount**”.

Finalmente, após validar o agrupamento podemos ver que, agora, temos apenas uma linha para cada indivíduo e o valor total que foi reivindicado.

Concluído o exercício, grave o ficheiro criado no “**Power BI**” com a designação “**exercicio\_1\_19\_R.pbix**”.

Ficheiro Home Transform Add Column View Tools Help						
Close & Apply	New Source	Recent Sources	Enter Data	Data source settings	Manage Parameters	Refresh Preview
Close	New Query	Sources	Data Sources	Parameters	Query	Advanced Editor
					Choose Columns	Remove Columns
					Keep Rows	Remove Rows
					Sort	Split Column
					Group By	Transform

= Table.TransformColumnTypes("#Promoted Headers",{{"Name", type text}, {"Branch", type text}, {"Date Claimed", type date}, {"Amount", type number}, {"Category", type text}}						
	Name	Branch	Date Claimed	Amount	Category	
1	Kian Thompson	Southampton	06/03/2013	472,24	Travel	
2	Millie Macdonald	Southampton	20/08/2016	16,51	Vehicle mileage	
3	Lewis Francis	Banbury	29/07/2015	87,84	Phone	
4	Sam Watson	Liverpool	06/07/2014	237,83	Accommodation	
5	Harriet Johnson	Banbury	07/06/2014	466,33	Miscellaneous	
6	Lewis Francis	Banbury	09/11/2015	20,58	Vehicle maintenance	
7	Henry Williams	Cardiff	06/06/2012	17,2	Online expenses	
8	Kian Thompson	Southampton	03/03/2013	183,45	Vehicle fuel	
9	Erin Dale	Lisburn	24/02/2012	147,8	Vehicle maintenance	
10	Joseph Fisher	Leeds	25/11/2015	196,52	Entertaining	
11	Lewis Francis	Banbury	30/06/2013	130,73	Office equipment	
12	Harriet Johnson	Banbury	23/07/2012	7,77	Research	
13	Lewis Francis	Banbury	22/06/2016	94,69	Vehicle fuel	
14	Bradley Wood	Liverpool	15/07/2012	178,97	Vehicle fuel	
15	Charlie Ross	Cardiff	29/11/2014	92,31	Accommodation	
16	Ethan Dean	Leeds	15/08/2013	51,69	Public transport	
17	Harrison Hart	London	08/06/2014	15,68	Stationery	
18	Benjamin Owen	Manchester	24/04/2015	183,6	Miscellaneous	
19	Jacob Spencer	Cardiff	31/07/2012	490,04	Accommodation	
20	Noah Johnson	London	01/11/2015	82,77	Research	
21	Ethan Wells	London	14/05/2012	134,74	Meals	

5 COLUMNS, 999+ ROWS Column profiling based on top 1000 rows