

Power Query

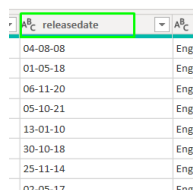
Uma das principais ferramentas para normalizar dados no Power BI é o Power Query. Dentro dele, é possível fazer diversas transformações como alterar nomes de colunas, preencher dados nulos, alterar tipos, criar colunas condicionais, enfim, preparar os dados que iremos utilizar em nossos datasets. Vamos explorar algumas dessas funcionalidades nas seções deste tópico.

Como exemplo, utilizaremos o dataset [audible](#), neste dataset temos um conjunto de dados sobre audiobooks. No *audible_uncleaned.csv* encontramos a versão não tratada desses dados, que serão um bom ponto de partida para nossos estudos!

Mudança de tipo de dados

Quando importamos dados para o Power BI utilizando uma fonte de dados que não carregue a informação de tipo (um csv por exemplo), há a possibilidade de que o Power BI faça a inferência do tipo incorreto de dado.

A **Figura 1** mostra a coluna release date que teve seu tipo inferido incorretamente como texto.



releasedate	
04-08-08	Eng
01-05-18	Eng
06-11-20	Eng
05-10-21	Eng
13-01-10	Eng
30-10-18	Eng
25-11-14	Eng
07-05-17	Eng

Figura 1 – Dado tipado incorretamente (Fonte da imagem: do autor)

Para transformar o tipo de uma coluna podemos clicar com o botão direito no nome dessa coluna no Power Query, e então selecionar a opção "Alterar tipo", definindo o tipo adequado, como exibido na **Figura 2**

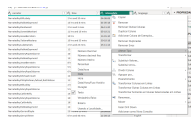


Figura 2 – Correção do tipo de dado (Fonte da imagem: do autor)

Fazendo isso, o Power Query adiciona a sequência de transformações a nova transformação desejada e já mostra como ficarão os dados após o comando "fechar e aplicar". Podemos ver o resultado na **Figura 3**

The screenshot shows the Windows Task Manager Performance tab. The CPU usage is highlighted with a green box and shows 100% usage. The 'Processes' tab is also visible on the right, showing a list of running processes.

Figura 3 – Tipagem de dado corrigida (Fonte da imagem: do autor)

Ajustando o formato dos dados

As vezes podemos ter um formato de input dos dados que não é interessante para uso em nossas visualizações. Na **Figura 4** podemos ver que tanto o autor quanto o narrador do audiobook trazem uma informação repetida e inadequada

[illegible]

Figura 4 – Colunas Autor e Narrador com dados ruidosos (Fonte da imagem: do autor)

Para resolver isso, podemos clicar com o botão direito na coluna e clicar em "Substituir Valores", o que invocará a caixa de diálogo da **Figura 5**, onde podemos preencher os trechos que desejamos substituir e as substituições que gostaríamos (para remover, basta deixar em branco).

Substituir Valores

Substitua um valor por outro nas colunas selecionadas.

Valor a ser substituído

Substituir por

☒ Aplicar imediatamente

☐ Gerar uma cópia dos dados antes de substituir.

☐ Substituir somente os valores numéricos.

Figura 5 – Dialogo de substituição de valores (Fonte da imagem: do autor)

Assim, é possível fazer a limpeza e verificar, na **Figura 6** que os dados estão mais adequados ao uso nas nossas visualizações

id	name	category	price	supplier
1	Mountain Dew 1.5 l	beverage	1.50	Mountain Dew Inc.
2	7-Up 1.5 l	beverage	1.50	7-Up Inc.
3	Orange Juice 1.5 l	beverage	2.50	Orange Juice Inc.
4	Orange Juice 0.5 l	beverage	0.75	Orange Juice Inc.
5	Apple Juice 1.5 l	beverage	2.50	Apple Juice Inc.
6	Apple Juice 0.5 l	beverage	0.75	Apple Juice Inc.
7	Orange Soda 1.5 l	beverage	2.50	Orange Soda Inc.
8	Orange Soda 0.5 l	beverage	0.75	Orange Soda Inc.
9	Apple Soda 1.5 l	beverage	2.50	Apple Soda Inc.
10	Apple Soda 0.5 l	beverage	0.75	Apple Soda Inc.
11	Mountain Dew 0.5 l	beverage	0.50	Mountain Dew Inc.
12	7-Up 0.5 l	beverage	0.50	7-Up Inc.
13	Orange Juice 0.2 l	beverage	0.25	Orange Juice Inc.
14	Apple Juice 0.2 l	beverage	0.25	Apple Juice Inc.
15	Orange Soda 0.2 l	beverage	0.25	Orange Soda Inc.
16	Apple Soda 0.2 l	beverage	0.25	Apple Soda Inc.
17	Mountain Dew 0.2 l	beverage	0.25	Mountain Dew Inc.
18	7-Up 0.2 l	beverage	0.25	7-Up Inc.
19	Orange Juice 0.1 l	beverage	0.125	Orange Juice Inc.
20	Apple Juice 0.1 l	beverage	0.125	Apple Juice Inc.
21	Orange Soda 0.1 l	beverage	0.125	Orange Soda Inc.
22	Apple Soda 0.1 l	beverage	0.125	Apple Soda Inc.
23	Mountain Dew 0.1 l	beverage	0.125	Mountain Dew Inc.
24	7-Up 0.1 l	beverage	0.125	7-Up Inc.
25	Orange Juice 0.05 l	beverage	0.0625	Orange Juice Inc.
26	Apple Juice 0.05 l	beverage	0.0625	Apple Juice Inc.
27	Orange Soda 0.05 l	beverage	0.0625	Orange Soda Inc.
28	Apple Soda 0.05 l	beverage	0.0625	Apple Soda Inc.
29	Mountain Dew 0.05 l	beverage	0.0625	Mountain Dew Inc.
30	7-Up 0.05 l	beverage	0.0625	7-Up Inc.

Figura 6 – Coluna Autor limpa (Fonte da imagem: do autor)

Separando informações agrupadas incorretamente

Na **Figura 7** podemos ver que há uma estruturação inadequada das informações na avaliação dos audiobooks, existem duas informações armazenadas: a avaliação média e também a quantidade de avaliações. Além disso, há um bloco de texto repetitivo que não se qualifica como informação para as nossas visualizações. Vamos resolver isso!

releaseDate	language	stars
04/08/2008	English	5 out of 5 star(s) ratings
01/05/2010	English	4.5 out of 5 star(s) ratings
06/11/2010	English	4.5 out of 5 star(s) ratings
09/03/2011	English	4.5 out of 5 star(s) ratings
10/12/2010	English	4.5 out of 5 star(s) ratings
06/03/2008	English	3 out of 5 star(s) ratings
29/11/2009	English	3 out of 5 star(s) ratings
02/09/2007	English	3 out of 5 star(s) ratings
05/05/2007	English	3 out of 5 star(s) ratings
24/08/2008	English	3 out of 5 star(s) ratings
29/01/2009	English	4 out of 5 star(s) ratings
06/08/2011	English	3 out of 5 star(s) ratings
21/09/2011	English	3 out of 5 star(s) ratings
09/03/2011	English	3 out of 5 star(s) ratings

Figura 7 – Dados de avaliação bem ruidosos (Fonte da imagem: do autor)

Primeiramente, podemos apagar o texto de Ratings com o que aprendemos na seção anterior. Em seguida, podemos utilizar a função de dividir coluna, como vemos na **Figura 8**.

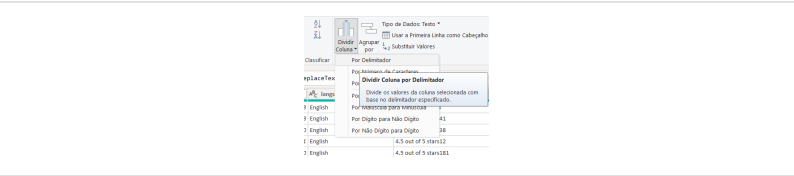


Figura 8 – *Dividir Coluna* (Fonte da imagem: do autor)

Por fim, podemos utilizar o restante do texto como um delimitador para separar as colunas como na **Figura 9**

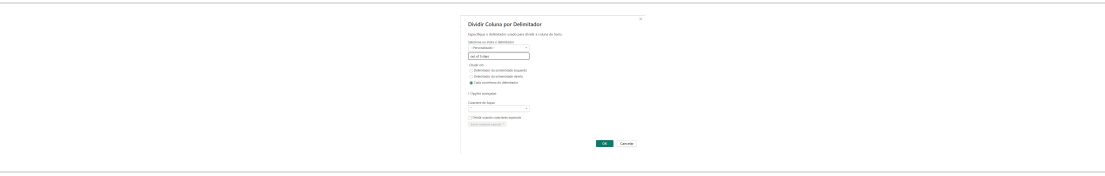


Figura 9 – *Utilizando delimitador personalizado* (Fonte da imagem: do autor)

Por fim, podemos renomear as colunas, clickando com o botão direito e escolhendo a opção renomear, para que os nomes se adequem melhor ao que elas representam, como na **Figura 10**

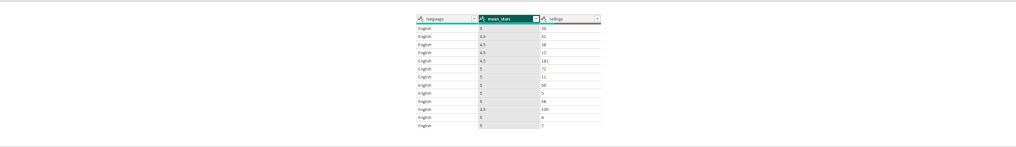


Figura 10 – *Nomes das colunas ajustados* (Fonte da imagem: do autor)

Com o Power Query conseguimos ferramental arrumar a bagunça que era essa base de dados, e assim poderemos gerar visualizações bem legais sobre esses dados!

No próximo tópico será mostrado como juntar os gráficos e construir um dashboard no Power BI.

Referências

<https://learn.microsoft.com/pt-br/power-query>