



## Power Query

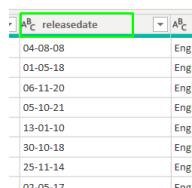
Uma das principais ferramentas para normalizar dados no Power BI é o Power Query. Dentro dele, é possível fazer diversas transformações como alterar nomes de colunas, preencher dados nulos, alterar tipos, criar colunas condicionais, enfim, preparar os dados que iremos utilizar em nossos datasets. Vamos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

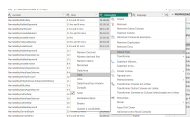
A **Figura 1** mostra a coluna release date que teve seu tipo inferido incorretamente como texto.



AB_c_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
02-05-17	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**

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

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).

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

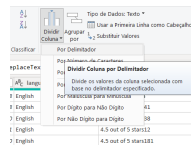
Figura 6 – Coluna Autor limpa (Fonte da imagem: do autor)

## 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!

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, clicando com o botão direito e escolhendo a opção renomear, para que os nomes se adequem melhor ao que elas representam, como na **Figura 10**

language	mean, stars	ratings
English	3	15
English	4.5	41
English	4.5	55
English	4.5	12
English	4.5	101
English	5	72
English	5	11
English	5	90
English	5	3
English	5	58
English	5.5	110
English	5	6
English	5	7

**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>