



POWER BI – MÓDULO I

SEMANA 04

Prof. Kaio Mesquita



Objetivos do curso

- *Instalar e configurar a ferramenta Power BI;*
- *Diferenciar e compreender conceitos de Business Intelligence, Big Data e Data Science;*
- *Compreender e avaliar problemas de maneira analítica, estratégica e fenomenológica;*
- *Criar tabelas de relacionamento;*
- *Apresentar resultados interativos;*
- *Tratar, Transformar e Carregar(ETL – Extract, Transform e Load) dados de forma otimizada;*
- *Analisar Séries Temporais;*
- *Criar Dashboard práticos e objetivos.*

Programa do curso

✓ 1. Primeiros passos com o Power BI (29/10)

✓ 2. Business Intelligence no Power BI (05/11)

✓ Modelagem, Relacionamento e DAX (12/11)

✓ 4. Limpeza, Transformação, Séries Temporais, Agregação e Filtros (19/11)

✓ Gráficos e Mapas (26/11)

✓ 6. Dashboards Interativos (03/12)

4.Limpeza,Transformação, Séries Temporais,Agregação e Filtro

- ✓ 4.1 Tarefa 01 - Verificação e Correção de Erros com Slice e Limpeza
- 4.2 Tarefa 02 - Limpeza, Transformação, Data analytics com P e R
- 4.3 Tarefa 03 - Quebra de dados e relacionamento
- 4.4 Tarefa 04 – Configurações regionais parciais por campo
- 4.5 Tarefa 05 – Limpeza e Transformações Profundas para criação de medidas
- 4.6 Tarefa 06 – Limpeza de dados null e atualização de dados em um arquivo
- 4.7 Tarefa 07 – Carregamento e Atualização de múltiplos arquivos

*Verificação

- ✓ Atualizações;
- ✓ Configurações Regionais

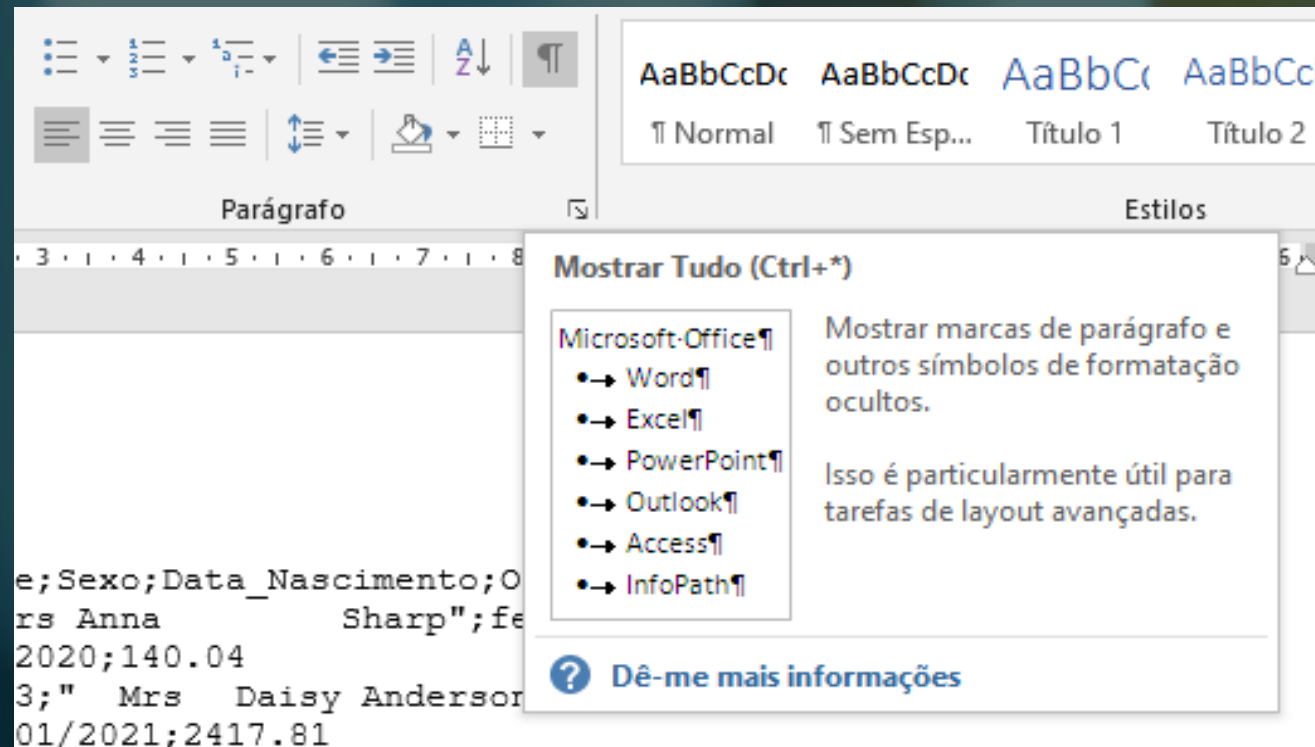
4.1 Tarefa 01 - Verificação e Correção de Erros com Slice e Limpeza

- ✓ Crie um ambiente de trabalho (pastas organizadas) para seus projetos e evite deixar espaços na nomenclatura.
- ✓ Altere a configuração **Global** para **Português (Brasil)** e a configuração **Local** para **Inglês (Estados Unidos)**.
- ✓ Abra o arquivo da **Tarefa 01.csv** em **abrir como no bloco de notas** e **verifique os possíveis erros**.

```
Username;Nome;Sexo;Data_Nascimento;Ocupação;Assinatura;Total_Gasto
Neave1959;"Mrs Anna      Sharp";feminino;02/02/1970;Engenheiro de Dados;18/
Nicterims1993;" Mrs    Daisy Anderson ";masculino;05/07/1985;Engenheiro de Da
Fouldlairity;"Ms Zoe    Peacock";masculino;03/06/1964;Analista Financeiro;04/12
Ancomettiody;" Mrs Libby    Williams ";feminino;04/09/1974;Estatístico;21/
```

4.1 Tarefa 01 - Verificação e Correção de Erros com Slice e Limpeza

- ✓ Agora abra um arquivo Word e arraste o arquivo excel para dentro dele.



4.1 Tarefa 01 - Verificação e Correção de Erros com Slice e Limpeza

- ✓ Utilize a ferramenta slice, depois a limpeza e se necessário substitua valores em branco.

The screenshot displays the Microsoft Power Query Editor interface. The 'Formato' (Format) dropdown menu is open for the 'Nome' column. The menu options include 'minúscula', 'MAIÚSCULA', 'Colocar Cada Palavra em Maiúscula', 'Cortar', 'Limpar', 'Adicionar', and 'Adicionar sufixo'. A tooltip for the 'Limpar' option is visible, stating: 'Remove os espaços em branco à esquerda e à direita de cada célula nas colunas selecionadas.' The background shows a data table with columns 'Username', 'Nome', and 'Ocupação'.

	Username	Nome	Ocupação
1	Neave1959	Mrs Anna Sharp	feminino
2	Nicterims1993	Mrs Daisy Anderson	masculino
3	Fouldlairity	Ms Zoe Peacock	masculino
4	Ancomettiody	Mrs Libby Williams	feminino
5	Onswis	Ms Katie Metcalfe	feminino
6	Majeappre	Mr Luca Vaughan	feminino
7	Theired	Mr Henry Haynes	masculino
8	Botter	Mr Noah Walton	feminino
9	Neaude1978	Mr Bailey Glover	masculino
10	Maress39	Mrs Isabella Fowler	feminino
11	Kehonell	Mrs Chloe Kirby	feminino
12	Thimay	Mrs Louise Clark	feminino

4.1 Tarefa 01 - Verificação e Correção de Erros com Slice e Limpeza

✓ Agora crie um DashBoard com as seguintes condições:

- Tabela com total de gastos por cliente
- Gráfico de barras com a média de gráficos por cliente (É necessário criar uma medida inicialmente)
- Gráfico de barras do sexo pelo total de gastos
- Todos os gráficos devem ter um fundo branco com títulos e subtítulos vermelhos
- Coloque uma logo e cabeçalho no seu projeto.

4.2 Tarefa 02 - Limpeza, Transformação, Data analytics com P e R

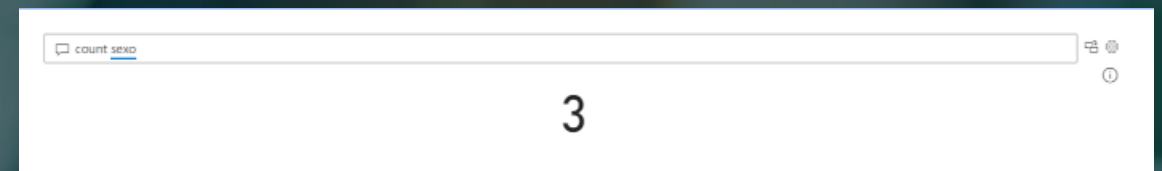
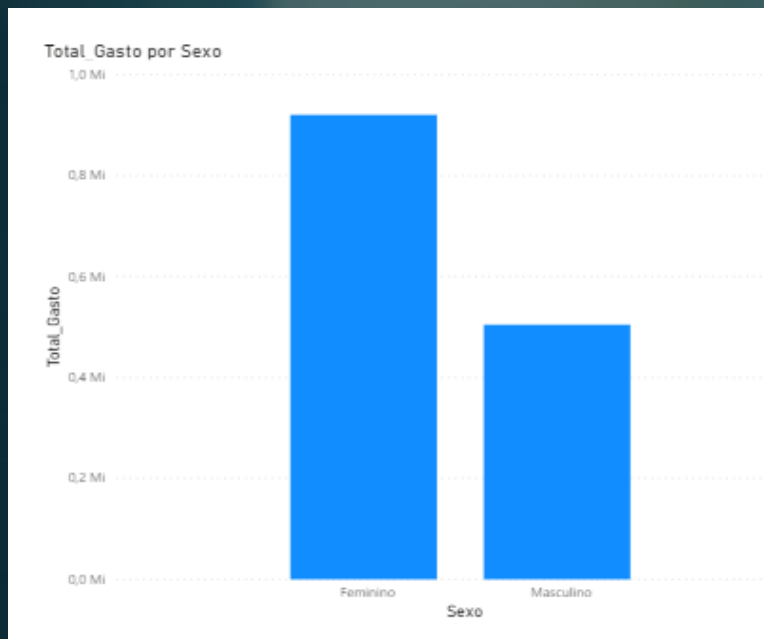
- ✓ Embora estejamos fazendo uma diferenciação entre limpeza e transformação dos dados, na prática é um único processo, sendo o processo de **data Reading**;
- ✓ Abra o arquivo da Tarefa 02 no bloco de notas;
- ✓ Corrija o encoding para UTF-8;
- ✓ No Power Query corrija os espaços em branco

```
compras - Bloco de Notas
Arquivo  Editar  Formatar  Exibir  Ajuda

Username;Nome;Sexo;Data_Nascimento;Ocupação;Assinatura;Total_Gasto
Neaude1978;" Mr          bailey Glover  ";masculino;01/12/1979;Analista de Intel.
Maress39;"Mrs    isabella  Fowler";feminino;04/08/1960;Arquiteto de Dados;24/0
Kehonell;" Mrs Chloe   Kirby  ";feminino;28/06/1949;Analista de RH;07/06/2020;
Thimpay;" Mrs louise   Clark   ";feminino;07/01/1995;Arquiteto de Dados;03/11/20
```

4.2 Tarefa 02 - Limpeza, Transformação, Data analytics com P e R

- ✓ Coloque comentários nas propriedades de etapas aplicadas (metadados);
- ✓ Crie um gráfico do gênero pelo total de vendas;
- ✓ Utilize a ferramenta P e R para verificar a quantidade de gêneros;
- ✓ Remova linhas em branco e verifique novamente a ferramenta.



4.2 Tarefa 02 - Limpeza, Transformação, Data analytics com P e R

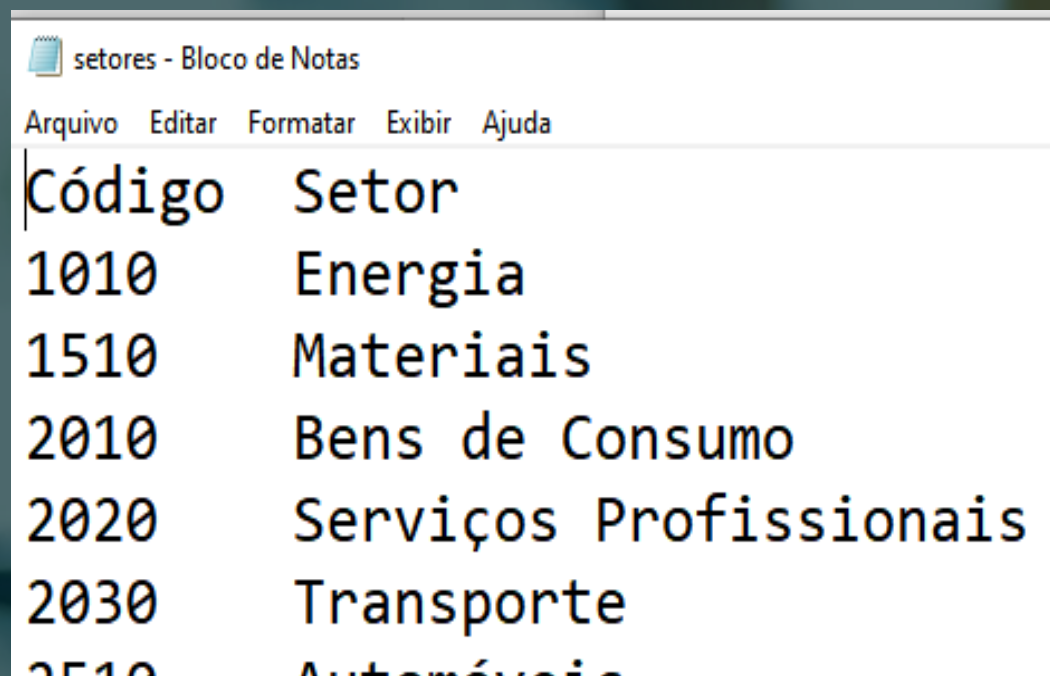
✓ Agora suponha que um gestor fez algumas perguntas, e você não colocou as informações que ele perguntou na sua apresentação. Utilize a ferramenta P e R para descobrir as seguintes respostas:

- Qual o total de pessoas cadastradas na nossa base?
- Qual a média do total de gastos?
- Qual o mínimo e máximo dos gasto?
- Qual a nossa última data de vendas?
- Quem são os usuários com gastos maiores que R\$ 175.000?

4.3 Quebra de dados e relacionamento

- ✓ Tente sempre limpar os dados primeiramente, antes de transformar, pois caso você faça o contrário, poderá estar transformando dados que não servirão para suas análises.

Depois carregue os arquivos no Power BI, comece pelo arquivo de **setor.csv** e corrija o **encoding**, além de verificar o tipo de delimitador como **tabulação**.



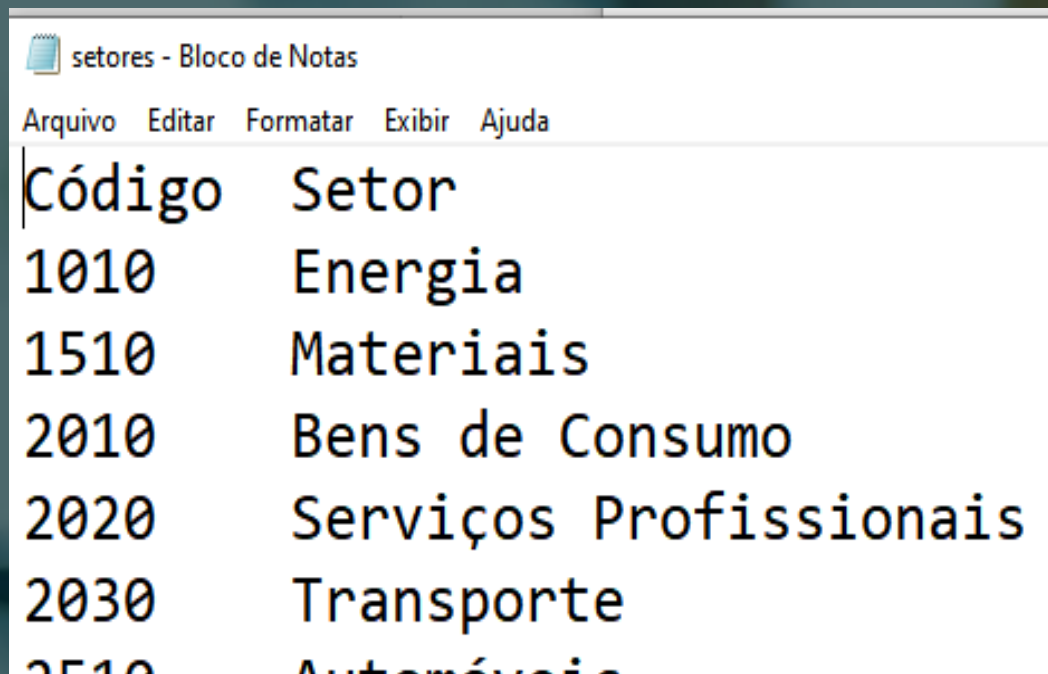
A screenshot of a Notepad window titled "setores - Bloco de Notas". The window has a menu bar with "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The text inside the window is as follows:

Código	Setor
1010	Energia
1510	Materiais
2010	Bens de Consumo
2020	Serviços Profissionais
2030	Transporte
2510	Automóveis

4.3 Quebra de dados e relacionamento

- ✓ Tente sempre limpar os dados primeiramente, antes de transformar, pois caso você faça o contrário, poderá estar transformando dados que não servirão para suas análises.

Depois carregue os arquivos no Power BI, comece pelo arquivo de **setor.csv** e corrija o **encoding**, além de verificar o tipo de delimitador como **tabulação**.

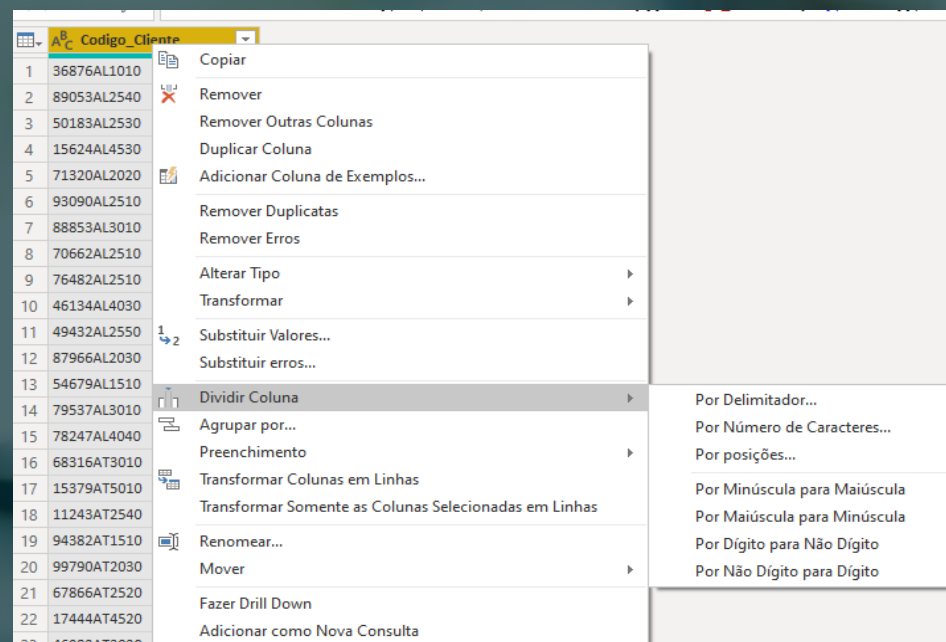


A screenshot of a Notepad window titled "setores - Bloco de Notas". The window has a menu bar with "Arquivo", "Editar", "Formatar", "Exibir", and "Ajuda". The text inside the window is as follows:

Código	Setor
1010	Energia
1510	Materiais
2010	Bens de Consumo
2020	Serviços Profissionais
2030	Transporte
2510	Automóveis

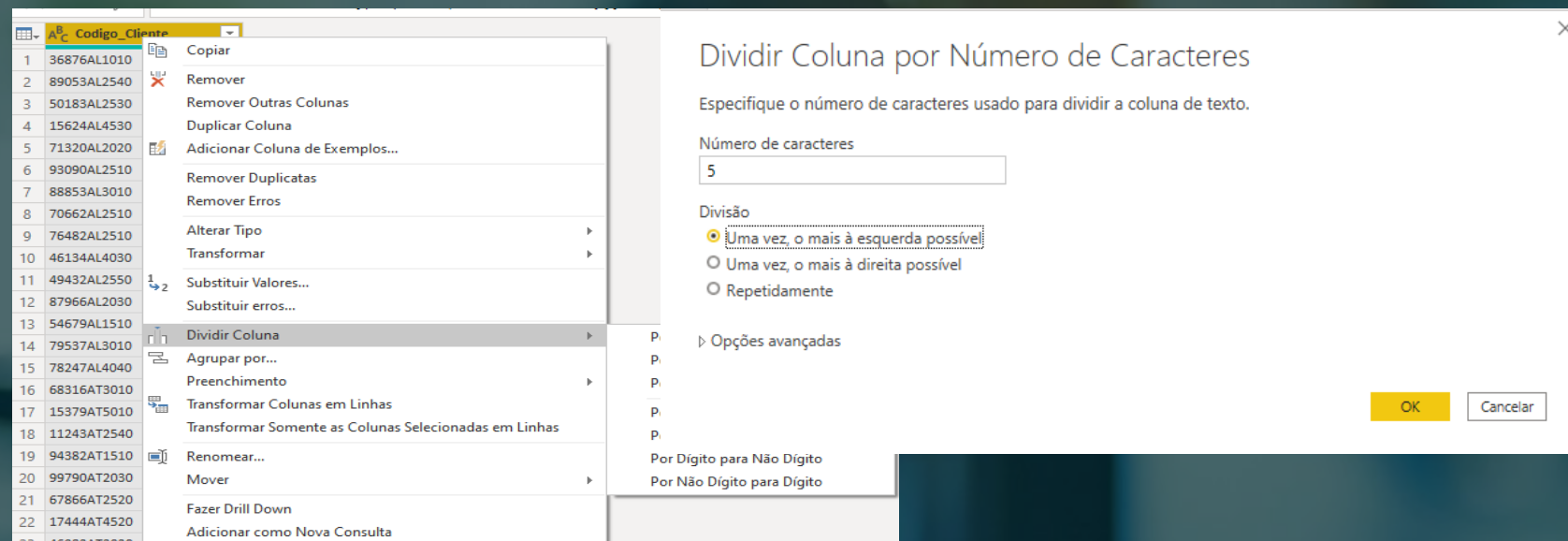
4.3 Quebra de dados e relacionamento

- ✓ Agora carregue o arquivo países.csv e se atente as devidas correções necessárias. Por fim carregue o último arquivo.
- ✓ Agora como processo de limpeza corrija o nome das colunas devidamente. Não há como decorar o processo, mas devemos aprender a corrigir quando necessário.
- ✓ **Clique na coluna com o botão direito, depois em dividir coluna e por números de caracteres.**



4.3 Quebra de dados e relacionamento

- ✓ Agora carregue o arquivo países.csv e se atente as devidas correções necessárias. Por fim carregue o último arquivo.
- ✓ Agora como processo de limpeza corrija o nome das colunas devidamente. Não há como decorar o processo, mas devemos aprender a corrigir quando necessário.
- ✓ **Clique na coluna com o botão direito, depois em dividir coluna e por números de caracteres, em 3 colunas diferentes.**

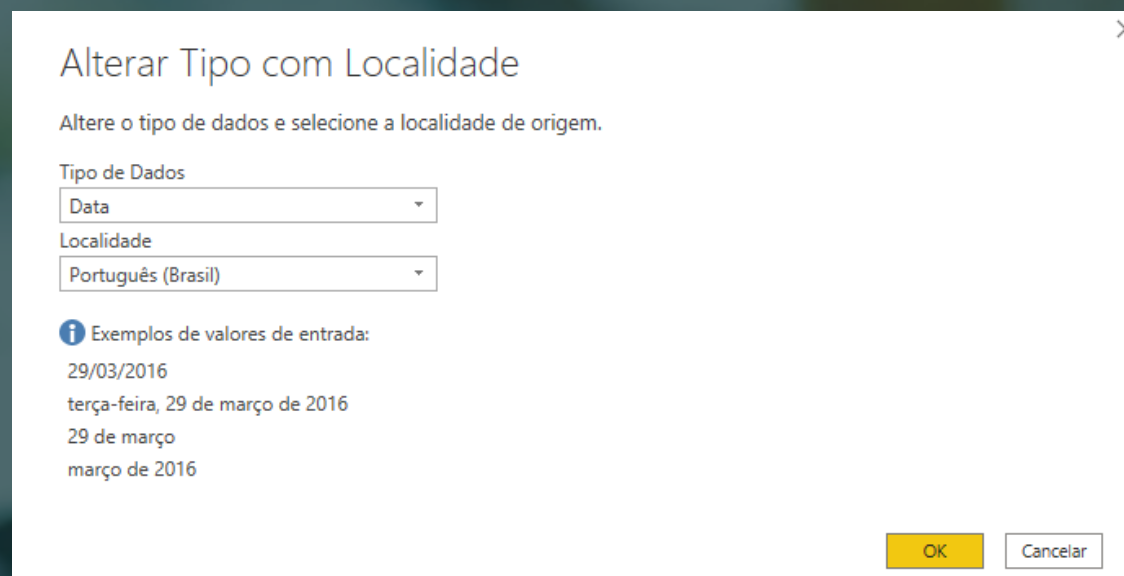
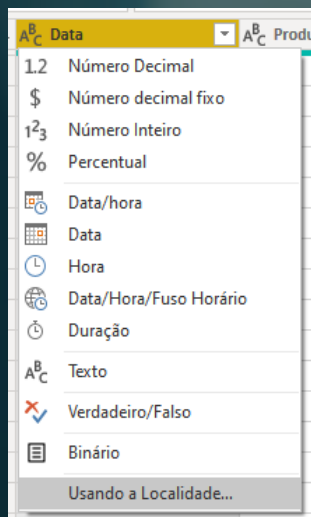


4.3 Quebra de dados e relacionamento

- ✓ • Agora crie o relacionamento entre as tabelas, verifique se a cardinalidade está correta para o problema.
- ✓ • Crie um Treemap do setor pela contagem do número de clientes
- ✓ • Verifique quais são os 5 setores com maior quantidade de clientes

4.4 Configurações regionais parciais por campo

- ✓ Carregue os dados da Tarefa 4, custos.csv e verifique as devidas correções;
- ✓ Perceba que a coluna data foi identificada como do tipo texto e não data. Na prática um tipo diferente de dados detém propriedades diferentes, então é necessário corrigir o tipo sempre.
- ✓ Por isso podemos corrigir a configuração regional em apenas uma coluna. Clique sobre o ícone do tipo e em seguida em Usando a Localidade.



4.4 Configurações regionais parciais por campo

✓ Agora aplique filtros de datas no menu de visualização para fazer gráficos com as seguintes características.

- Valor de vendas para dias inferiores ao dia 20/03/2021
- Valor de vendas entre as datas 15/03/2021 e 25/03/2021
- Preço de Custo para datas maiores que 08/03/2021
- Ajuste o drill-Down para visualização por dia e por ano para visualizar as diferenças

4.5 Limpeza e Transformações Profundas para criação de medidas

- ✓ Abra o arquivo da tarefa 5 e verifique o tipo de tabulação.

Arquivo Editar Formatar Exibir Ajuda		
Despesas 2021 XYZ Corporation		
Categoria	Data	Quantidade
Acomodação	12/04/2021	R\$ 247,20
	01/02/2021	R\$ 34,62
	08/07/2021	R\$ 367,89
	28/01/2021	R\$ 79,76
	07/02/2021	R\$ 263,57
	08/11/2021	R\$ 4,75

4.5 Limpeza e Transformações Profundas para criação de medidas

- ✓ Carregue os dados no Power BI, modifique o **encoding para UTF-8**. E perceba que ele detectou 4 colunas e algumas estão em branco.

despesas_empresa.txt

Origem do Arquivo: 65001: Unicode (UTF-8) | Delimitador: Tabulação | Detecção de Tipo de Dados: Com base nas primeiras 200 linhas

Column1	Column2	Column3	Column4
Despesas 2021 XYZ Corporation			
Categoria	Data		Quantidade
Acomodação	12/04/2021	R\$ 247,20	
	01/02/2021	R\$ 34,62	
	08/07/2021	R\$ 367,89	
	28/01/2021	R\$ 79,76	
	07/02/2021	R\$ 263,57	
	08/11/2021	R\$ 4,75	
	01/12/2021	R\$ 150,26	
Taxas Bancárias	01/09/2021	R\$ 37,00	
	09/01/2021	R\$ 10,78	
	08/11/2021	R\$ 101,33	
	10/05/2021	R\$ 64,96	

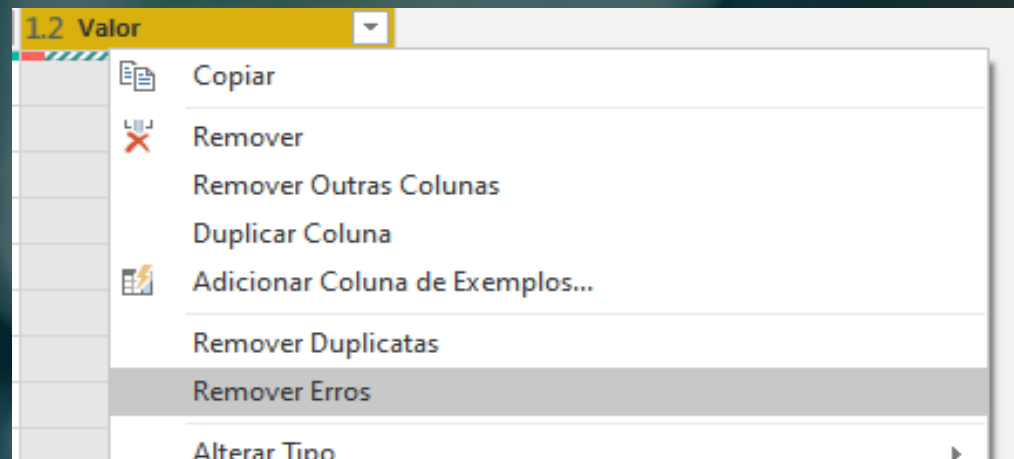
4.5 Limpeza e Transformações Profundas para criação de medidas

- ✓ Retire as primeiras linhas do dataset.

Agora coloque a primeira linha como cabeçalho. Porém a terceira coluna está sem título e a quarta coluna está vazia. Apague a quarta coluna e renomeei a terceira.

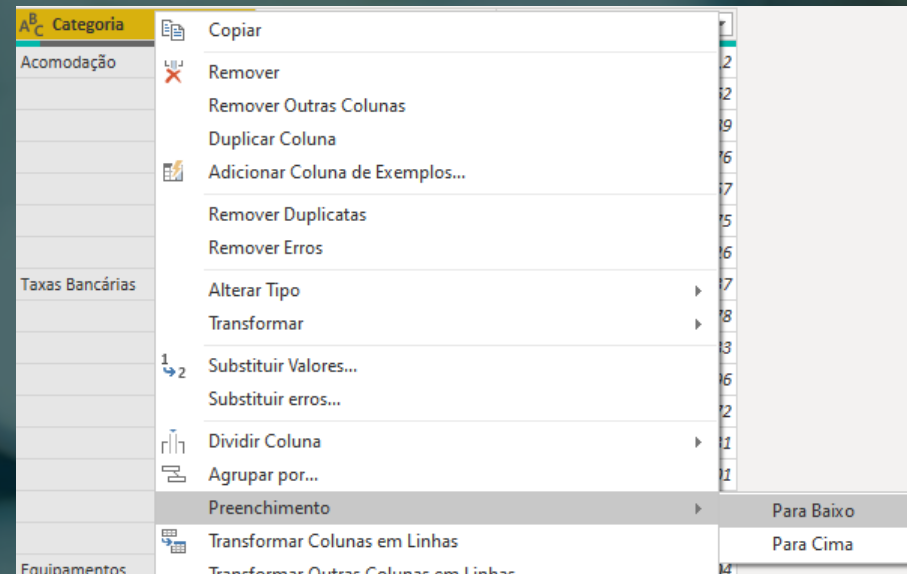
Clique com o botão direito e clique em remover erros.

*Outra possibilidade é calcular a média dos valores corretos e substituir nas linhas que têm erros, respeitando o teorema do limite central.



4.5 Limpeza e Transformações Profundas para criação de medidas

- ✓ Na primeira coluna temos valores em branco, que para uma modelagem preditiva seria um enorme problema, por isso também devemos corrigir.
- ✓ Clique com o botão direito sobre a coluna e clique em substituir valores, na primeira coluna mantenha um espaço em branco e na segunda digite null;
- ✓ Agora clique em preenchimento e para baixo, e os dados existentes preencherão os valores null.



4.5 Limpeza e Transformações Profundas para criação de medidas

✓ Agora crie as seguintes medidas e as utilize para visualização por tipo de categoria.

- Média por categoria
- Mediana por categoria
- Desvio Padrão Amostral por categoria
- Coeficiente de Variação por categoria
- Utilize o P e R para verificar as suas visualizações

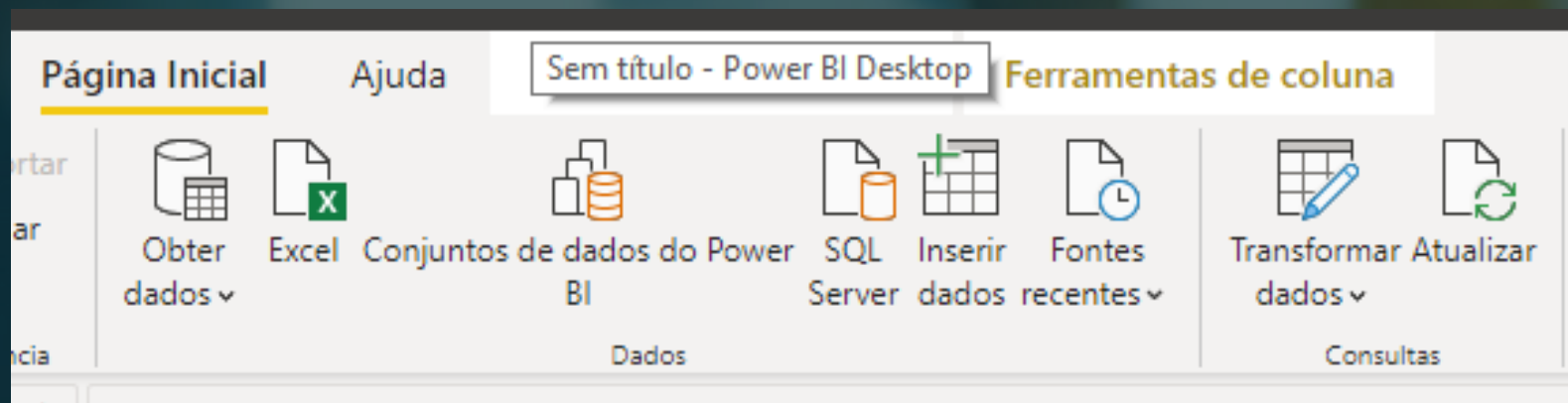
4.6 Limpeza de dados null e atualização de dados em um arquivo

- ✓ Agora vamos utilizar os dados referentes a Tarefa6, cadastro.xls. Carregue os dados. Verifique que existem as colunas finais estão vazias.
- ✓ Selecione todas as colunas em branco e exclua as colunas.
- ✓ Agora selecione a coluna título e arraste ela para o início do dataset. Reorganize as colunas de uma forma que facilite o seu trabalho.

Tipo Sanguineo	Peso	Altura	Column16	Column17	Column18	Column19	Column20
A+	80	178					
A+	90	175					
A+	88	189					
A+	67	182					
B+	73	183					
B-	78	178					
A-	90	179					
B+	81	172					
O+	73	181					
B+	69	183					

4.6 Limpeza de dados null e atualização de dados em um arquivo

- ✓ Posteriormente abra o arquivo excel coloque uma nova linha de dados e feche o arquivo. Agora vá no menu iniciar e clique em atualizar e veja essa nova linha sendo inserida.



4.6 Limpeza de dados null e atualização de dados em um arquivo

✓ Vamos para a seguinte atividade:

- Crie 30 novas linhas no excel com dados fictícios e atualize-os na sua planilha
- Transforme a coluna altura para metros (crie uma nova coluna ou faça o cálculo direto na fórmula)
- Agora crie uma coluna que calcule o IMC com as informações do peso e altura e apague essas duas colunas. Pesquise a fórmula antes.

$$[\text{Peso}] / (([\text{Altura}] / 100) * ([\text{Altura}] / 100))$$

A empresa está interessada em saber quais as profissões que estão sendo mais prejudicadas em relação a saúde dos seus funcionários. Crie uma medida de média do IMC e uma visualização com as profissões pela média, e de acordo com a classificação abaixo indique quais setores estão mais prejudicados.

4.6 Limpeza de dados null e atualização de dados em um arquivo

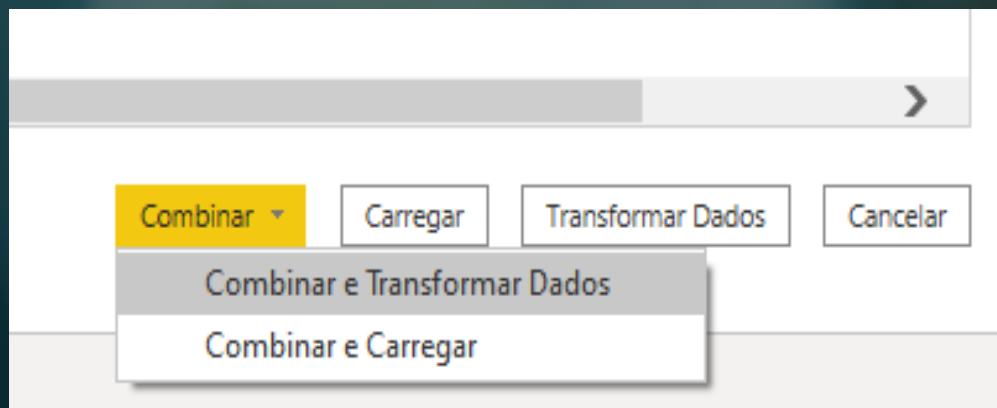
Classificação do IMC

- Abaixo do peso: **IMC** abaixo de 18,5.
- Peso normal: **IMC** entre 18,5 e 24,9.
- Sobrepeso: **IMC** entre 25 e 29,9.
- Obesidade Grau I: **IMC** entre 30 e 34,9.
- Obesidade Grau II: **IMC** entre 35 e 39,9.
- Obesidade Grau III: **IMC** acima de 40.

4.7 Carregamento e Atualização de múltiplos arquivos

- ✓ Agora importe o arquivo da Tarefa7. Mas visualize que existem mais de um arquivo csv na pasta. Então escolha o tipo pasta no Power BI para importar todos de uma vez (**Não coloque o arquivo ceará.csv na pasta**).

Serão identificados todos os arquivos da pasta. Então clique em combinar e transformar. E você terá acesso a um arquivo de cada vez para transformação. Verifique se existe alguma erro nos arquivos, corrija-os e posteriormente carregue-os.



4.7 Carregamento e Atualização de múltiplos arquivos

- ✓ Abra o Power Query e perceba que foi criado uma nova coluna contendo identificador do arquivo de onde vieram os dados.

Para que isso funcione é necessário que as colunas sejam iguais em todos os arquivos importados com os mesmo cabeçalhos, embora tenham dados diferentes. Perceba que os dados já foram agregados e qualquer aplicação realizada conterà as informações de todos os arquivos.



The screenshot shows the Power Query Editor interface. At the top, there is a formula bar with a dropdown menu showing 'Table.TransformColumns'. Below the formula bar, there is a table with two columns: 'A' and 'C'. The header row is highlighted in yellow and contains the text 'Nome da Origem'. The table contains 12 rows of data, all with the value 'vitoria.csv' in column 'C'.

	A	C
12		vitoria.csv
13		vitoria.csv
14		vitoria.csv
15		vitoria.csv
16		vitoria.csv
17		vitoria.csv
18		vitoria.csv
19		vitoria.csv
20		vitoria.csv
21		vitoria.csv
22		vitoria.csv
23		vitoria.csv
24		vitoria.csv

4.7 Carregamento e Atualização de múltiplos arquivos



Faça as seguintes atividades:

- Crie uma visualização com estado e valor de venda
- Coloque o arquivo do estado do Ceará e atualize o Power BI
- Verifique a visualização recém criada se houve alteração
- Crie uma visualização do valor de venda por data, exceto o ano de 2015
- Crie um gráfico de pizza do valor de venda por fabricante, filtrado apenas para o Rio de Janeiro e São Paulo
- Importe um novo tema para o seu projeto

Projeto

- ✓ Estamos no ano de 2018 e você trabalha numa empresa que presta consultoria de Engenharia de Tráfego. A Prefeitura de Fortaleza está decidindo se vai ou não fazer intervenções viárias na Rua Alberto Magno, devido algumas reclamações recebidas sobre número alto de acidentes naquela região.

Sua empresa foi contratada para fazer uma apresentação e ajudar a prefeitura a tomar uma decisão se é cabível realmente fazer alguma intervenção (Binário, semaforização vertical e horizontal, equipamentos de controle de fluxo, etc). Lhe disponibilizaram um conjunto de dados de acidentes dos anos de 2016 e 2017 de algumas ruas das proximidades, para que trabalhe com uma série temporal.

Faça as devidas limpezas e tratamentos, monte uma apresentação objetiva com as informações e medidas que julgar importante e por fim dê uma conclusão sobre o caso.

OBRIGADO!



Kaio Mesquita



kaio@det.ufc.br



kaio_mesquitaa