Power BI

# O que significa a sigla “BI”?

BI é a sigla para Business Intelligence, traduzindo para o português, Inteligência do Negócio, mas qual é o conceito do BI?

# O conceito de BI

O conceito é relacionado ao trabalho relacionado aos dados da empresa, trazendo métricas para que os responsáveis pelas decisões da empresa possam dar os direcionamentos necessários para o negócio baseado nos dados coletados.

Por incrível que pareça esse conceito surgiu na década de 60, mas na época a tecnologia era muito limitada e ficou atrelado às decisões estratégicas.

Hoje em dia, tudo o que está a nossa volta pode gerar algum dado que pode ser usado, por exemplo, em uma estatística. Isso para as empresas é algo muito precioso, pois com dados, elas podem definir critérios para entender o que está acontecendo e definir o rumo do negócio.

Em outras palavras, as empresas podem tomar decisões baseadas em fatos, tendo como base os dados e não mais no achismo.

Na maioria das vezes a pessoa que vai receber esses dados não é alguém técnico, por isso é importante escrever, detalhar e organizar as análises e decisões tomadas, para simplificar, contar a história da análise ou Storytelling.

Para apresentar essa história de maneira fácil para consulta, são feitos os Dashboards, ou painéis, traduzindo para o português.

Nos Dashboards tem elementos visuais como gráficos que facilitam e muito a interpretação dos dados e melhor, ele pode ser atualizado em tempo real, assim possibilitando a análise constante sobre os diferentes setores de uma empresa.

# Visão geral do Power BI

É importante entender os vários elementos que compõem a plataforma Power BI para criar soluções efetivas.

Conjuntos de dados, relatórios e dashboards, são elementos que são organizados em um espaço chamado **Workspace** e são criados em capacidades específicas.

O Power BI oferece diversas capacidades, como capacidades compartilhadas e dedicadas para hospedar e entregar conteúdo. O **Workspace** citado antes, são contêineres que organizam os elementos do BI, possibilitando a colaboração e compartilhamento de conteúdo entre colegas.

Os conjuntos de dados permitem a **importação e conexão** de diferentes fontes de dados, assim os tornando disponíveis para os membros de um **workspace**, garantindo a consistência e confiabilidade dos dados.

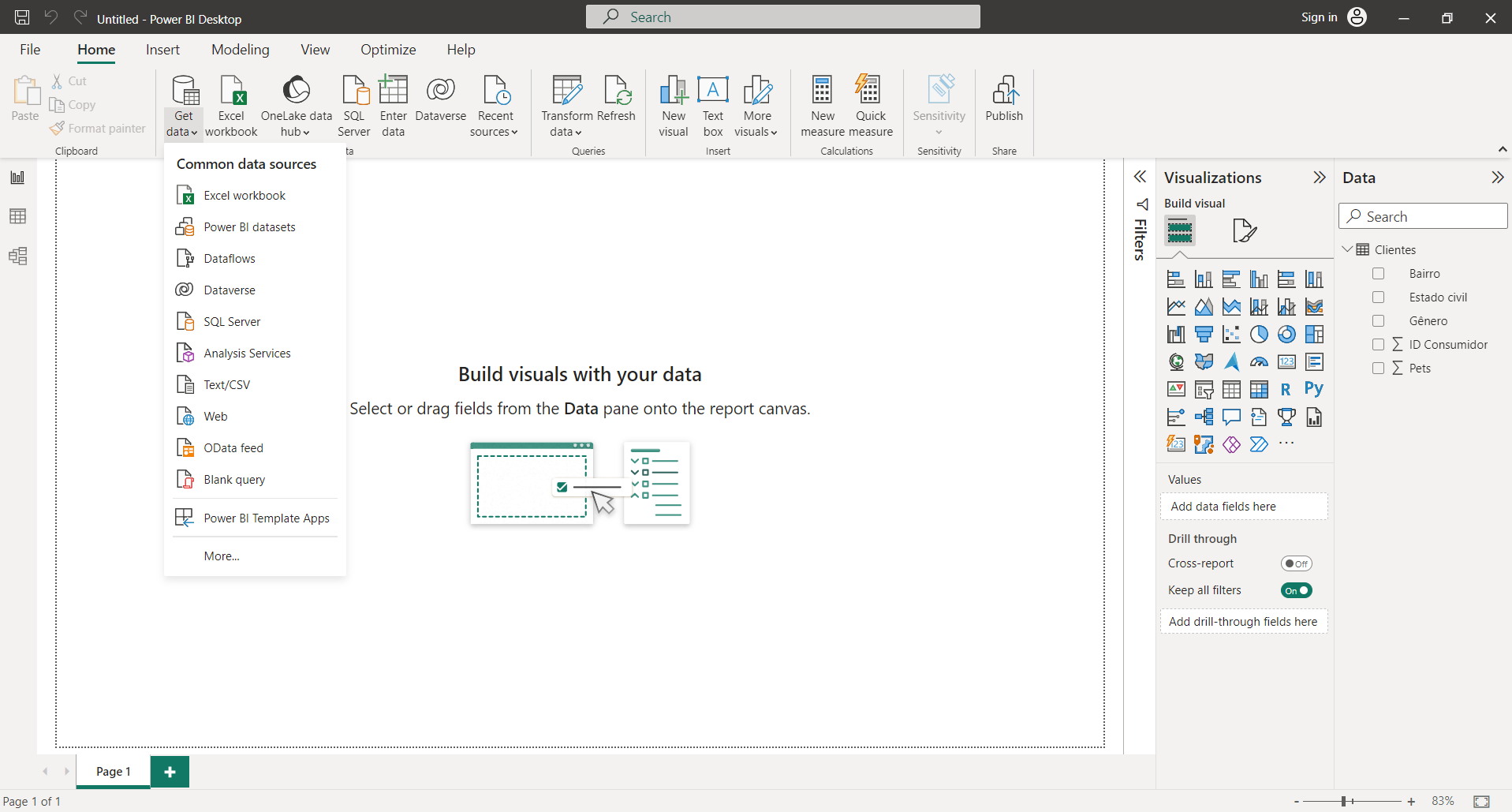
## Relatórios no Power BI

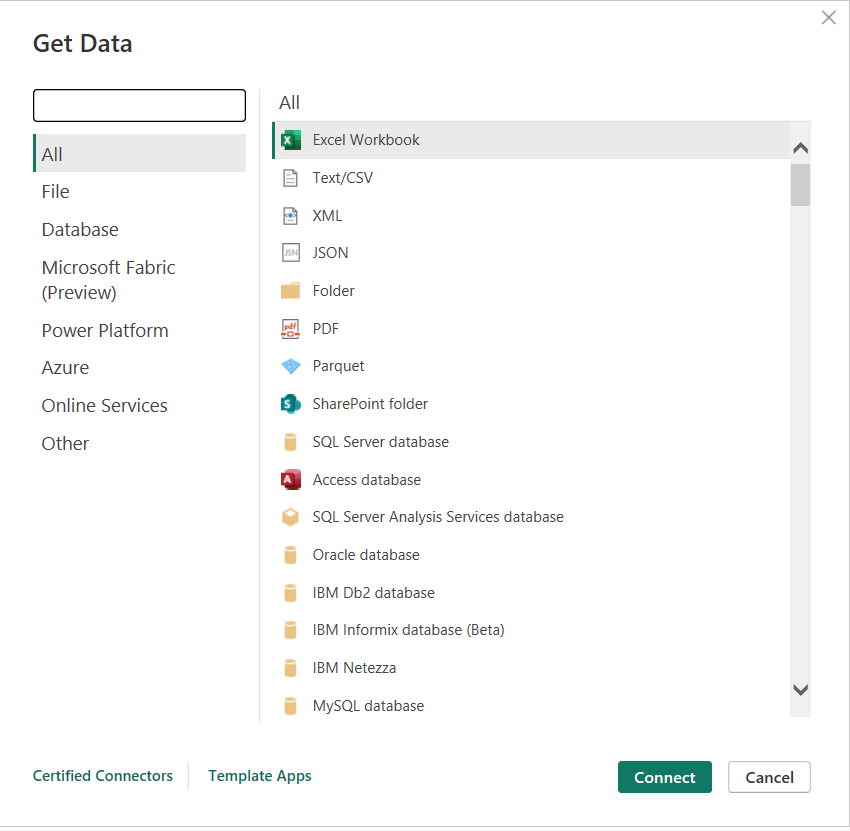
Os **relatórios** no Power BI consistem em **visualizações**, como gráficos e mapas, podendo ser criados do zero ou importados de outros usuários. Existem modos de exibição de leitura e edição, permitindo que os usuários interajam ou modifiquem o conteúdo, conforme suas permissões.

## Dashboards

Já os **dashboards** são telas que apresentam **blocos e widgets**, oferecendo uma visão geral e resumida das informações cruciais para a tomada de decisões, além de poderem ser personalizados e compartilhados com outros.

## Importando dados no Power BI





Dentro do Power BI podemos importar vários tipos de dados, como por exemplo, arquivos Text (txt), arquivos CSV, planilhas do Excel, importar dados do SQL Server, JSON, XML, arquivos do Microsoft Access, Azure e vários outros.

Depois que escolher o formato desejado clique em “Connect” e você pode começar a organizar os seus preciosos dados.

## Importando dados da Web

Na situação de importar uma planilha Google Sheets temos duas opções:

* Uma é procurando a opção específica **“Google Sheets”** depois de clicar em **“Get Data”**.

* A outra opção é a conexão **Web**.

Ambas as opções permitem importar dados de planilhas Google Sheets, no entanto, há diferenças entre as duas opções e isso vou explicar agora.

### Conexão Web

Essa é uma opção que permite importar dados de uma planilha Google Sheets para o Power BI através de um link compartilhado da planilha. Depois de escolhida, você precisa da **URL** da planilha para conectá-la com o Power BI, abaixo vou deixar alguns avisos sobre essa opção de conexão web:

* **Link Compartilhado:** Para usar a conexão via web, é necessário ter o link compartilhado da planilha no Google Drive, em outras palavras, a planilha precisa deve ser compartilhada e acessível a partir do link fornecido.

* **Atualização Manual:** Se os dados na planilha forem atualizados, você precisa atualizar a conexão no Power BI na mão para conseguir os dados mais recentes. Para fazer isso você clica no botão direito no conjunto de dados no Power BI e selecionando "**Refresh Data**".
* **Segurança Limitada:** É importante considerar as implicações de segurança ao compartilhar uma planilha. Se alguém além de você tiver acesso ao link compartilhado, pode ver tudo o que tem na planilha, e isso vai por mim, pode ser bem desconfortável.

### Google Sheets

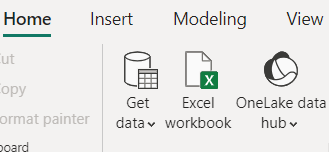
Como já dito, no começo desse tópico, existe uma opção específica com o nome Google Sheets que é fornecida pelo próprio Power BI para importar de forma direta as planilhas do Google, sem a necessidade de um link de compartilhamento, abaixo estão umas características dessa opção:

* **Conexão Direta:** Com a opção Google Sheets você se conecta diretamente à sua conta do Google e seleciona a planilha que você deseja, você pode ver que isso elimina a necessidade de criar e compartilhar links e depender de conexões web.

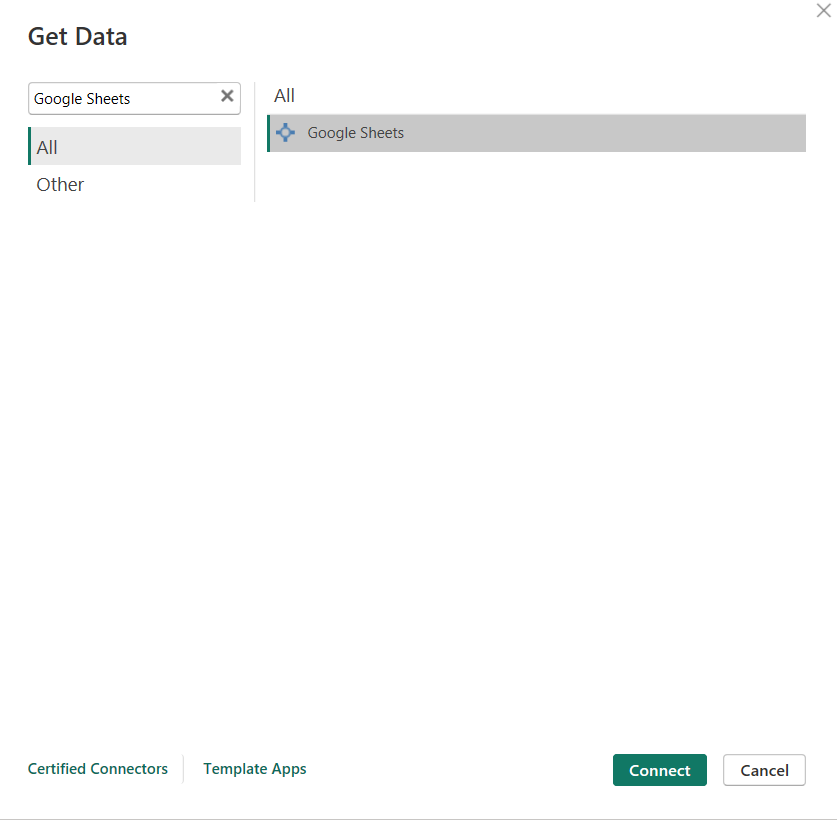
* **Melhor Segurança:** Com essa opção, você garante uma camada adicional de segurança, pelo motivo de não precisar compartilhar um link da planilha. Os dados são importados da sua conta do Google, mantendo a confidencialidade dos dados.

Caso queira usar, siga o tutorial:

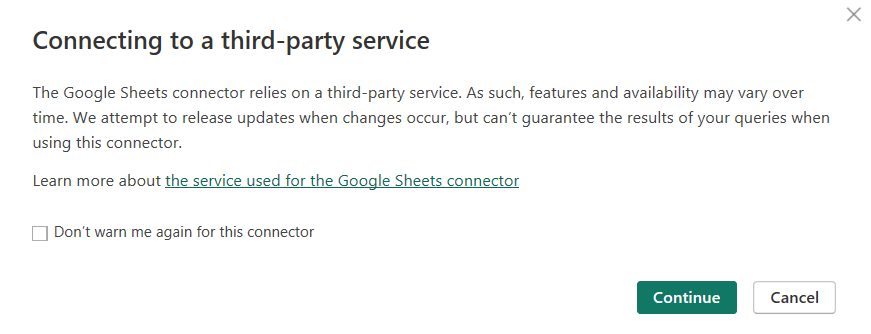
**1° -** Clique em **Get Data**:



Depois que a aparecer uma janelinha, pesquise por **“Google Sheets”** ou **“Planilhas Google”**, clique na opção e em seguida em **“Connect”**.



Depois de clicar em Connect, irá aparecer uma janela com informações, clique em **“Continue”**.



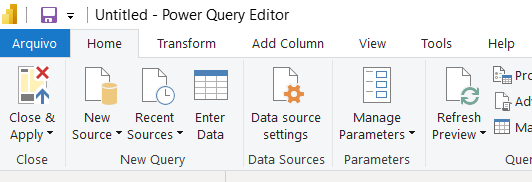
Agora basta copiar a URL da planilha e colar e em seguida clique em **Ok**.

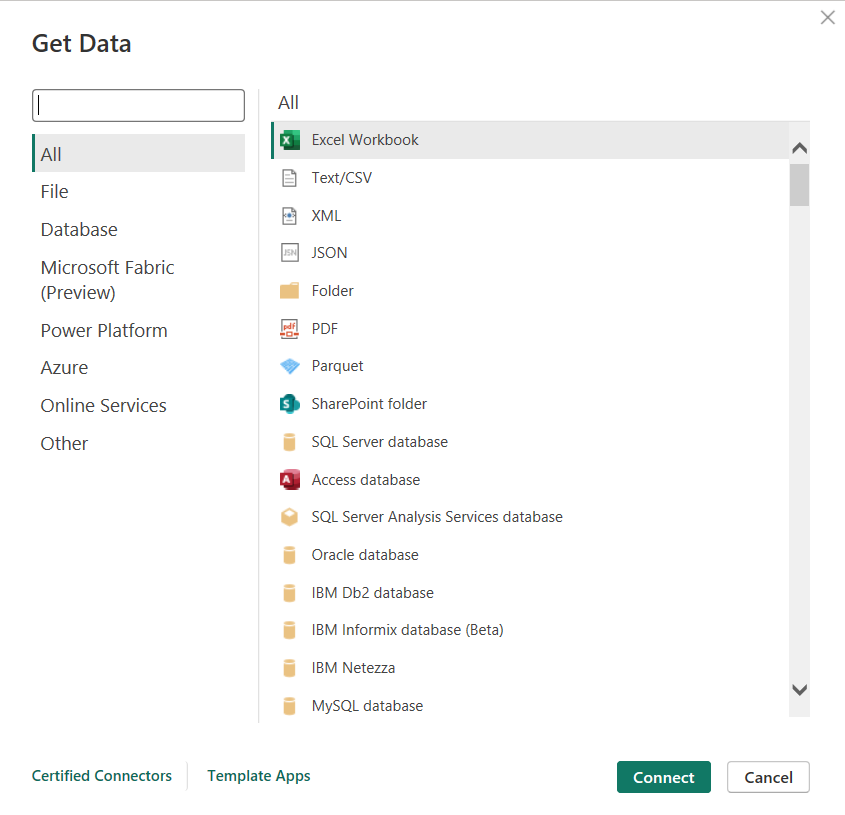


Agora você pode começar a tratar os seus dados.

## Importando do Excel

Dentro do próprio Power Query Editor você pode importar mais fontes de dados e para isso você clica em **“New Source”**, fica ao lado do **“Close & Apply”**. Você vai perceber que vai abrir a mesma janela onde você escolhe a fonte que deseja importar.





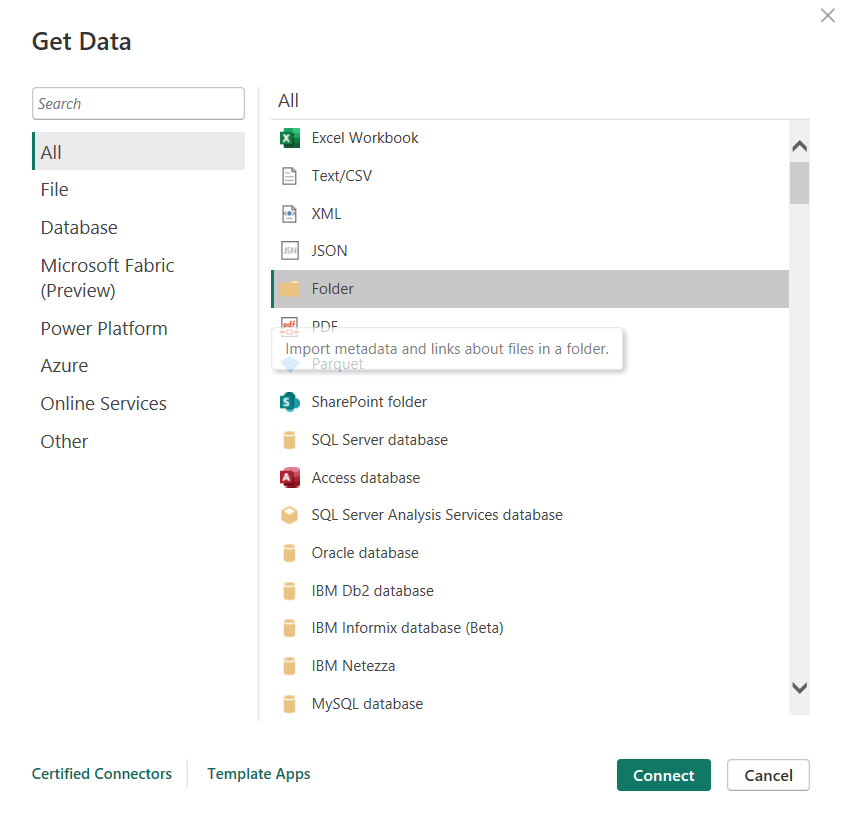
Caso os dados que for importar sejam do Excel, escolha a opção **“Excel Workbook”** e em seguida clique em **“Connect”**, depois procure o seu arquivo e trate da maneira mais adequada para você.

Só tem um problema, nós não podemos importar mais de um arquivo Excel, para evitar dor de cabeça e ficar importando um arquivo de cada vez, podemos fazer a seguinte solução:

* Pegar os arquivos Excel e guarda-los em uma pasta;
* No Power Query, depois de clicar em **“New Source”**, em vez de clicar em **“Excel Workbook**”, escolha a opção “**Folder”**;
* Clique em **“Connect”** e em seguida informe o diretório da pasta;
* Pronto! Você importou uma pasta com mais de um arquivo Excel.

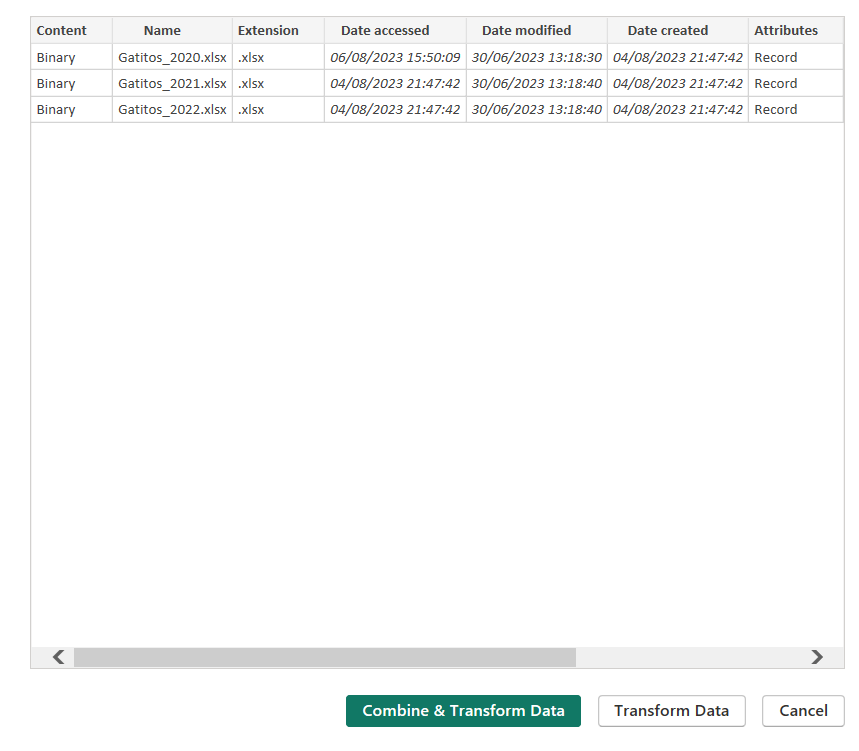
Lembrando que para isso você tem que saber muito bem onde a pasta está guardada na sua máquina.

Para os leigos, o formato de um arquivo Excel é **“.xlsx”.**





Aqui é onde você coloca o diretório/caminho da pasta onde estão os arquivos.



Em seguida vai aparecer essa janela contendo os arquivos que o Power BI encontrou baseado no diretório/caminho que foi informado, aqui no caso ele encontrou três arquivos Excel, ou seja, três arquivos .xlsx.

Uma coisa incrível é que clicando em **“Combine & Transform Data”** ele vai transformar essas três planilhas em uma só, e esse é o propósito, trazer e combinar mais de um arquivo Excel.

Vimos a diferença em importar um único arquivo Excel e importar vários arquivos de uma vez.

## Mais bases de dados

Existem vários sites que disponibilizam bases de dados gratuitas, especialmente destinadas a estudantes, como eu, que desejam praticar e explorar dados reais.

Um dos sites recomendados é o [Open Data NI](https://www.opendatani.gov.uk/). Ele oferece uma grande variedade de base de dados que são organizados por áreas. O site é em inglês, caso você queira testar seus conhecimentos na língua, use o site normalmente. Para os que não querem, é possível usar o recurso de tradução do navegador para facilitar o entendimento.

Outro site interessante é o [Dataset Search](https://datasetsearch.research.google.com/), que é disponibilizado pelo Google. Através dele, você pode pesquisar e encontrar conjuntos de dados relevantes para a sua área de interesse.

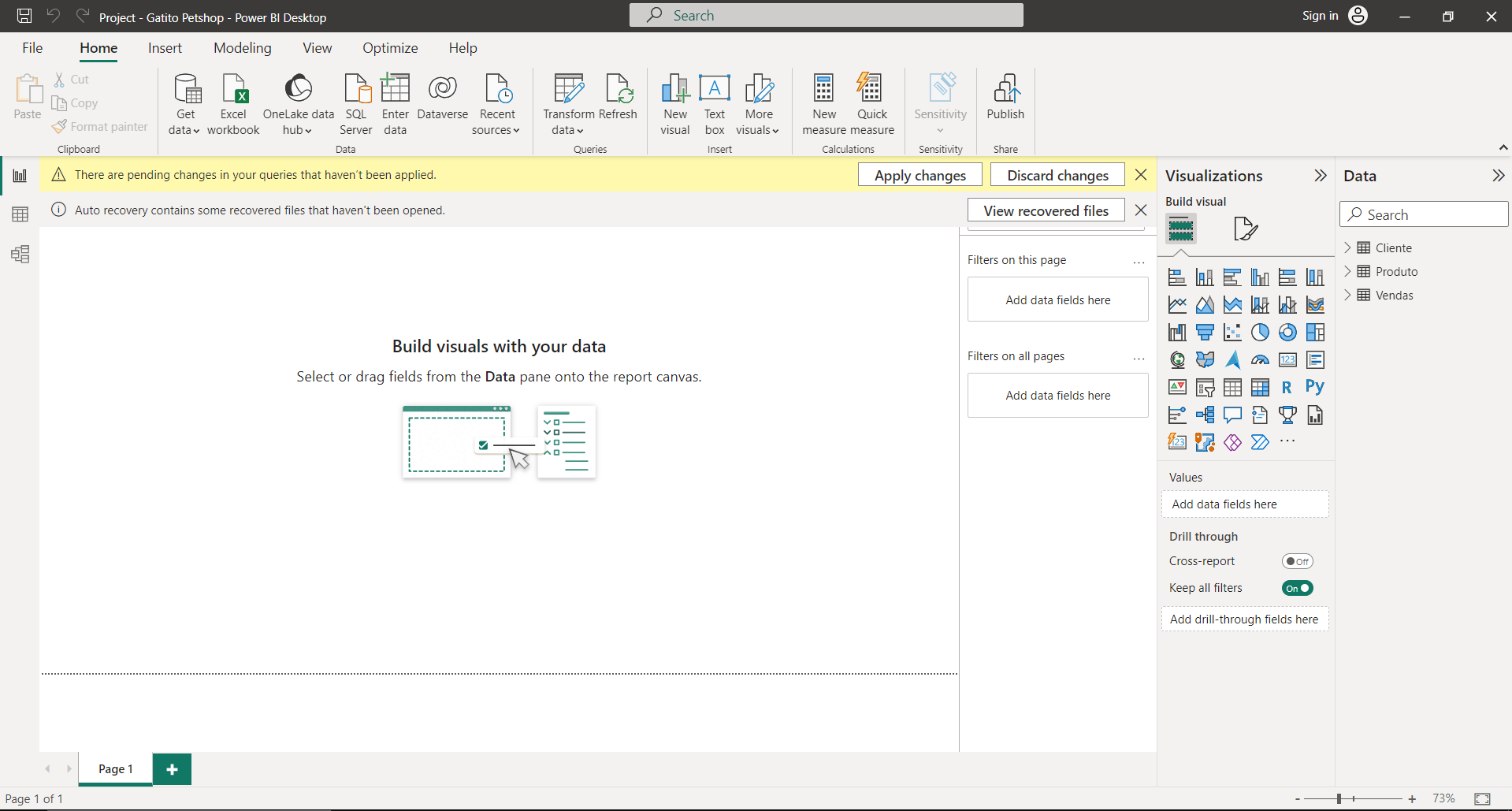
Para quem quer explorar dados brasileiros tem o [Instituto Brasileiro de Geografia e Estatística (IBGE)](https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html). No site oficial da instituição tem vários datasets disponíveis e eles são compostos por diversos levantamentos estatísticos realizados pelo instituto, proporcionando assim uma grande quantidade de informações para estudo.

Além do IBGE tem o [Portal Brasileiro de Dados Abertos](https://dados.gov.br/dados/conjuntos-dados), que fornece um grande conjunto de dados do governo.

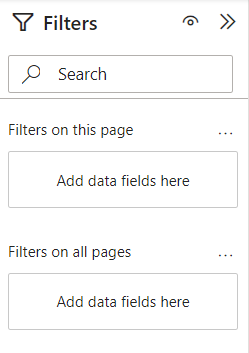
Saindo do Brasil e indo para a Microsoft, ela oferece uso do [banco de dados AdventureWorks](https://docs.microsoft.com/pt-br/sql/samples/adventureworks-install-configure?view=sql-server-ver15&tabs=ssms). Ele pode ser usado em conjunto com outros softwares da área de dados, deixando possível várias opções de exploração e prática.

Por último e não menos importante, tem o [Kaggle](https://www.kaggle.com/datasets) é um dos sites mais conhecidos na área de Data Science. Utilizado por estudantes e profissionais para compartilhar dados interessantes dos mais diversos tipos. Estudar pelo Kaggle pode revelar uma infinidade de possibilidades para aprimorar suas habilidades de análise de dados.

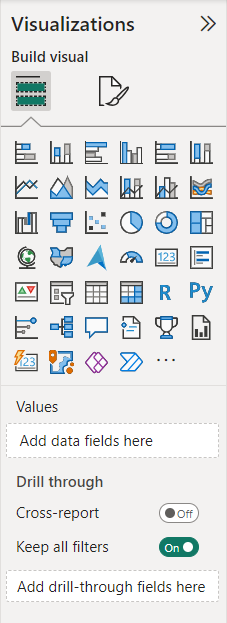
## Conhecendo um pouco sobre as áreas do Power BI



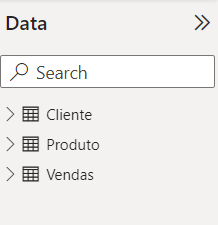
Nessa parte destacada em vermelho é onde vamos criar e desenvolver o nosso dashboard com os nossos dados, vamos chama-lo de **Canvas**.



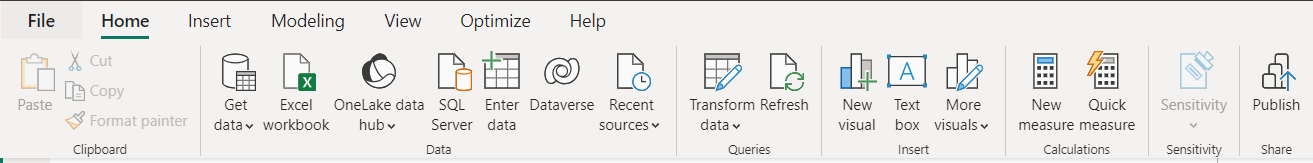
Ao lado direito do Canvas temos os **Filtros/Filters**, onde podemos trabalhar com seleções para modificar os nossos visuais.



Ao lado dos Filtros, temos as **Visualizações/Visualizations**, aqui temos todos os tipos de gráfico, visuais e muito mais coisas que vou conhecendo conforme vou estudando.



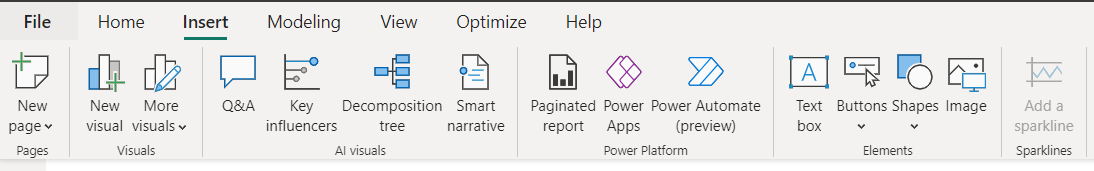
Temos a aba **Dados/Data**, é onde ficam as tabelas e colunas que importamos para dentro do Power BI.



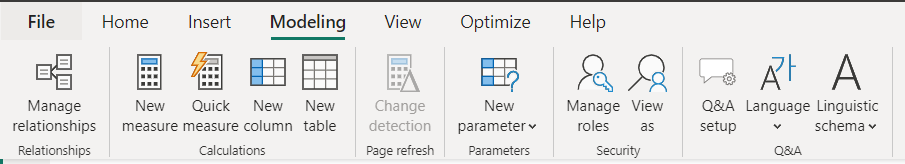
Aqui na parte superior na aba **Home**, vimos funções como o **Get Data** que permite a importação de dados, o **Excel Workbook**, onde importamos dados do Excel e o **Transform Data**, que nos permite acessar diretamente o Power Query.

Sim! O **Power Query** não é acessado somente quando queremos importar alguma base para o BI, podemos acessá-lo à vontade.

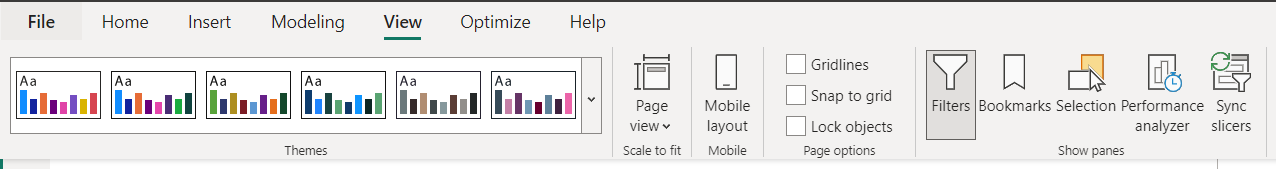
As áreas novas que eu for conhecendo, vou procurar documentar aqui para enriquecer o conhecimento que eu for adquirindo, claro você que estiver lendo também.



Na aba **Insert**, temos um recurso que tem no Word, Power Point e Excel, que é poder inserir uma **Text Box,** **Buttons**, **Shapes** e **Images** para os nossos dashboards.

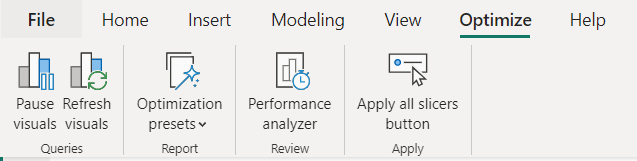


Na parte de **Modeling**, o nome já é bem auto explicativo, é onde podemos acrescentar coisas para as nossas modelagens, como usar o **New Measure** para criar novas medidas para o dashboard, criar novas colunas usando o **New Column** e o **New Table** para criar novas tabelas quando for necessário.



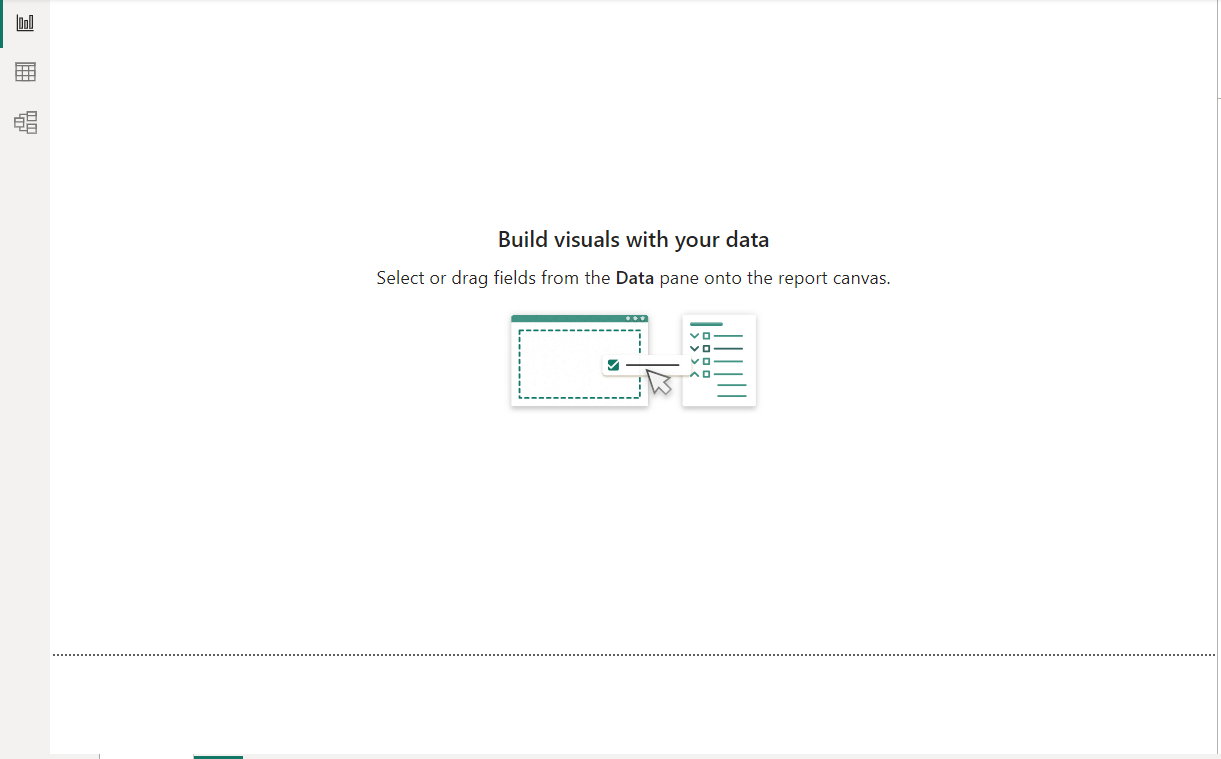
Na aba **View**, podemos observar que o Power Bi já traz para nós layouts prontos com vários jogos de cores, e isso podemos personalizar para ficar da maneira que a gente quer.

Também podemos trabalhar com **Filtros/Filters**, filtros antes, **Bookmarks** e Seleções usando o **Selection**.

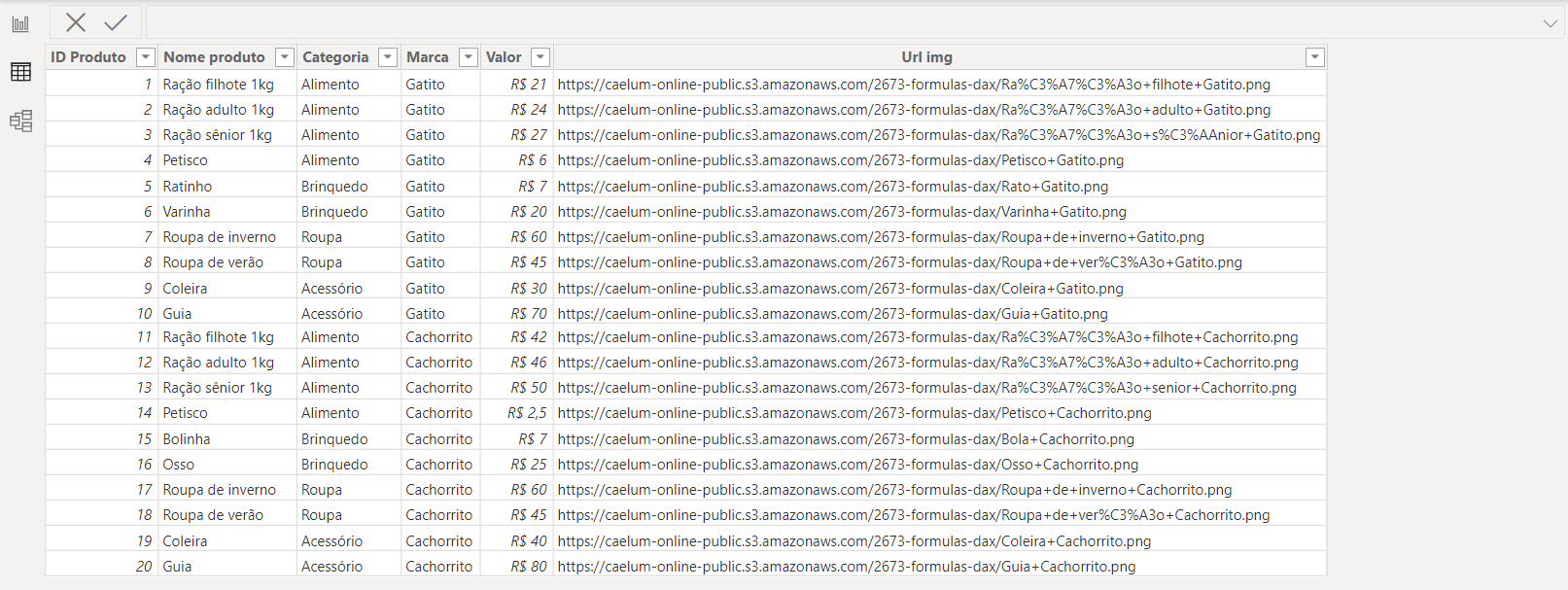


Temos a guia **Optimize**, onde aqui podemos trabalhar com a performance dos dados e modelos.

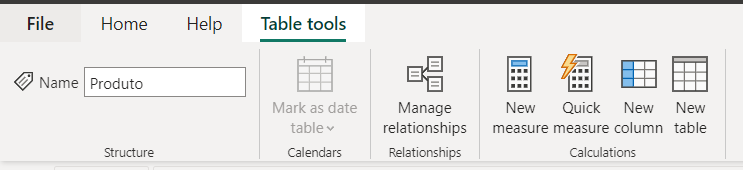
E relembrando, tudo o que for aprendendo será relatado aqui.



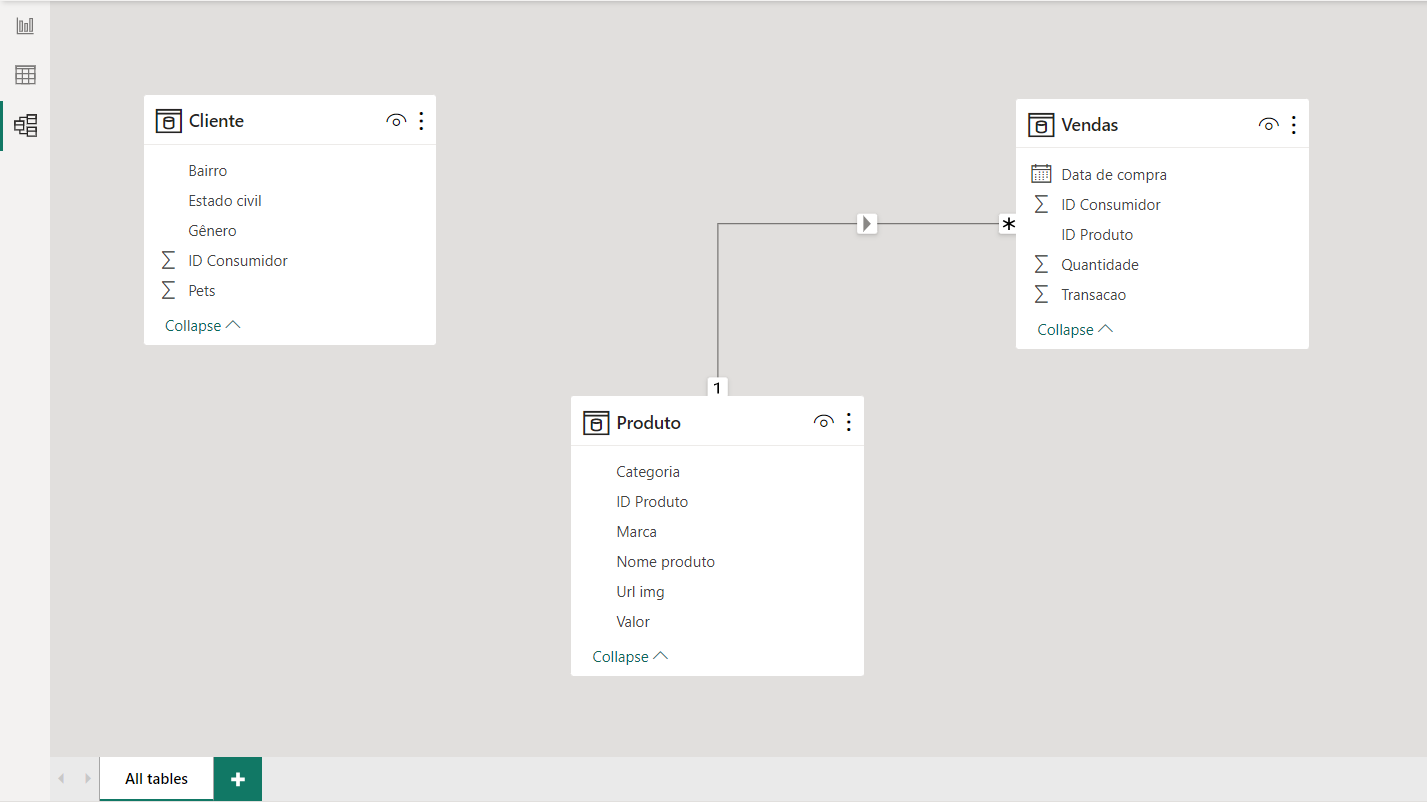
Você está vendo esses três botões em destaque no lado esquerdo do Canvas? O primeiro deles já está selecionado que é a tela do próprio Canvas que é chamado de **Report View** ou **Exibição de Relatório** caso o seu BI esteja em português.



Essa tela é a **Data View** ou **Exibição de Dados**, é o segundo botão que tem forma de planilha do Excel. Nela podemos ver as nossas tabelas e as colunas com os nossos dados.



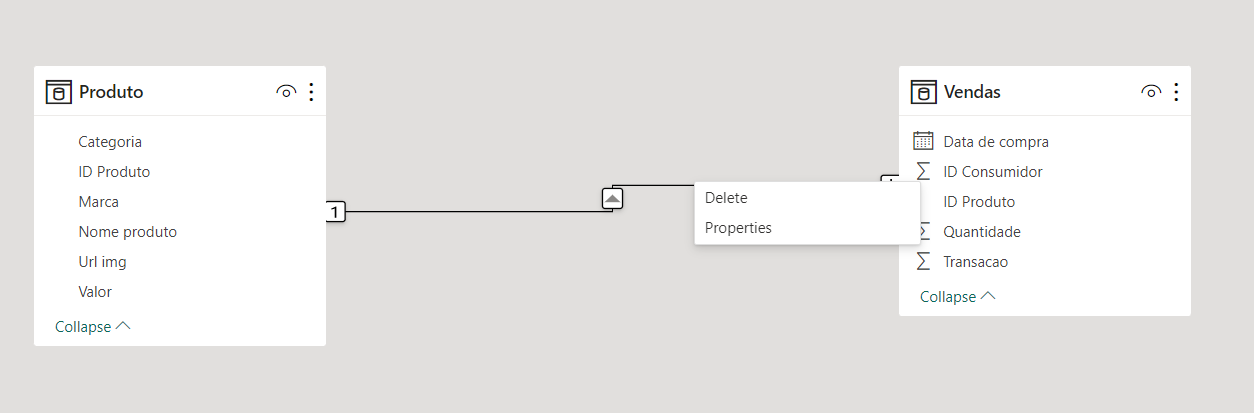
Também podemos manipular os dados, criar novas colunas e realizar operações entre os dados das colunas, isso usando a linguagem DAX.

  
Por último temos o **Model View**/**Exibição do Modelo**, nela podemos ver as tabelas presentes no nosso modelo, também podemos enxergar os relacionamentos presentes entre as entidades, bem tranquilo para quem já estudou modelagem de dados.

## Relacionamentos das Tabelas

Como você viu no Model View, mostra as tabelas e seus respectivos relacionamentos, mas eu quero ver mais sobre a relação das tabelas, como eu faço isso?

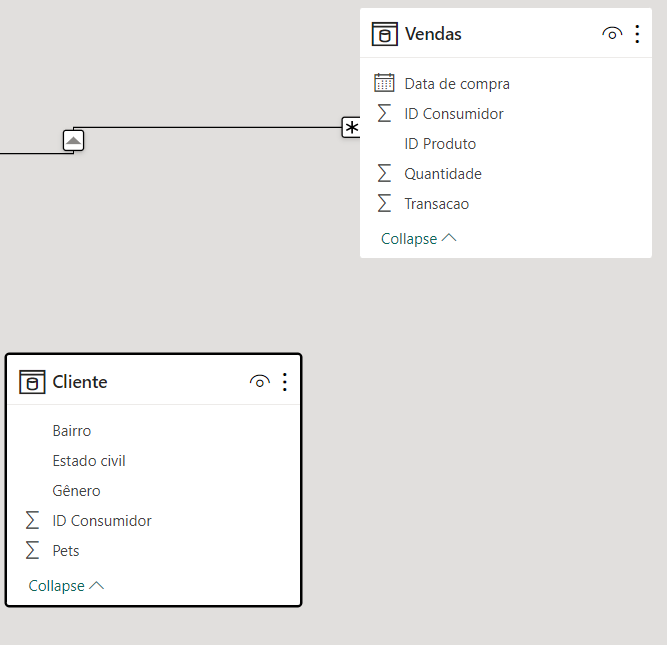
* Primeiro vá com o botão direto em cima da linha do relacionamento e clique em **Properties**;
* Depois vai aparecer uma tela com os detalhes sobre o relacionamento entre as tabelas e você vai perceber também que o Power BI tem a capacidade de identificar relacionamento entre entidades.



## 

Ok. Mas existem tabelas que tem um campo em comum que o Power BI não identificou como uma relação, o que eu faço?

A resposta é simples, vamos fazer nós mesmos!



Vê essas duas tabelas, elas têm um campo em comum e temos que relacioná-las. O campo em questão é o **ID Consumidor**, para criar o relacionamento, selecione o campo da tabela pai e arraste para a tabela filho, ou seja, selecione o ID Consumidor da **tabela Cliente** e arraste para a **tabela Vendas**.

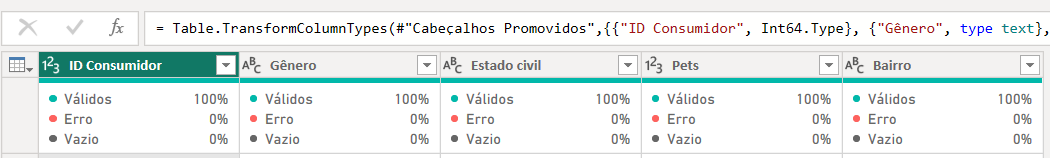
Caso não haja nenhum problema, o relacionamento será criado sem erros e você pode checar em **Properties**, mas caso um problema apareça irá mostrar uma aba mostrando onde está o conflito.

## Qualidade da Coluna

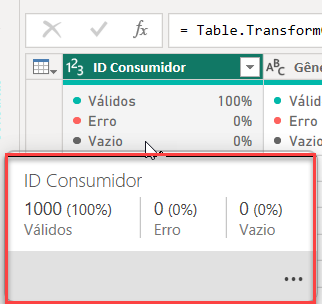
É um recurso do BI que rotula os valores em linhas em cinco categorias, fornecendo informações sobre a qualidade dos dados em cada coluna:

* **Válido (verde):** Indica que os valores na coluna estão corretos e dentro dos critérios definidos.
* **Erro (vermelho):** Mostra a presença de erros na coluna, indicando que os valores não estão de acordo com as regras ou critérios estabelecidos.
* **Vazio (cinza escuro):** representa valores ausentes ou nulos na coluna, indicando que não há dados presentes.
* **Desconhecido (verde pontilhado):** indica a presença de erros em uma coluna, resultando em uma qualidade de dados desconhecida para os demais valores.
* **Erro inesperado (vermelho pontilhado):** identifica a ocorrência de erros inesperados na coluna, que não se enquadram nas categorias anteriores.

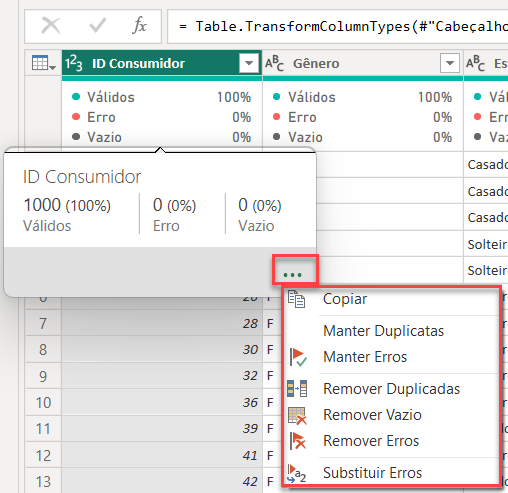
Esses indicadores são exibidos abaixo do nome da coluna. O número de registros em cada categoria de qualidade de coluna é apresentado em porcentagem, veja a imagem abaixo:



Ao passar o mouse sobre qualquer uma das colunas, você consegue ver a distribuição numérica da qualidade dos valores em toda a coluna:



Ao selecionar o botão de reticências (...), são exibidos botões de ação rápida que permitem realizar operações nos valores:



Esse recurso de Qualidade da Coluna do Power BI permite uma visão rápida e clara sobre a qualidade dos dados em cada coluna.

Entretanto, devemos ficar atentos a um quesito muito importante quando se trata do Power BI, que é o fato de se tratar de uma ferramenta que tem como padrão resumir os dados, principalmente por questão de performance.

Pensando nisso, o BI oferece a opção de filtrar os dados das tabelas, por meio de duas opções: criação de perfil da coluna com base nas primeiras 1000 linhas, que é a opção padrão; e criação de perfil da coluna com base em todo o conjunto de dados.

Essa opção de filtragem por todo o conjunto pode ser especialmente importante para garantir uma análise mais precisa e abrangente dos seus dados.

Você consegue encontrar essa opção na parte inferior do Power Query.

Com a opção de criação de perfil da coluna com base em todo o conjunto de dados ativada, o Power BI vai verificar todas as linhas do conjunto de dados, permitindo identificar padrões, distribuições e problemas de qualidade que podem não ser detectados apenas com uma amostra limitada de linhas.