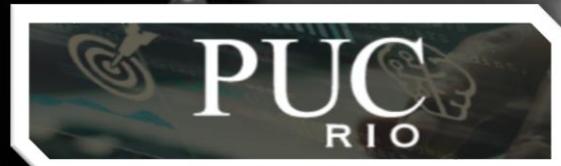


# ENGENHARIA DE DADOS

MVP

Michele Tempos Sarai | [micheletempos@gmail.com](mailto:micheletempos@gmail.com)

 <https://www.linkedin.com/in/michele-tempos-sarai-99497510b/>



# RESUMO

Este projeto visa criar um pipeline de dados completo, abrangendo pesquisa, coleta, modelagem, ingestão, categorização e análise de dados.

É fundamental escolher uma solução em nuvem que ofereça escalabilidade e agilidade para lidar com volumes variáveis de dados e permitir a análise eficiente. Isso garantirá uma tomada de decisão informada e a capacidade de integrar tecnologias avançadas para análises mais profundas.

A escolha da solução em nuvem e a construção do pipeline são partes cruciais deste projeto, capacitando a empresa a prosperar em um ambiente de negócios dinâmico e orientado por dados.

# Sumário

RESUMO .....	2
1. Solução Escolhida .....	4
2. Objetivo .....	6
3. Escolha da Base de dados .....	7
4. Coleta .....	8
4.1 Restauração do banco de dados .....	8
5. Modelagem e catálogo de dados .....	11
5.1 Modelagem.....	12
5.2 Catalogando os dados.....	13
5.2.1 Mapa Unificado dos Dados.....	14
5.2.1 Catálogo de Dados.....	17
6. Carga.....	22
6.1 WorkSpace.....	22
6.1.1 Visualizando as opções do Fabric .....	22
6.1.2 O Data Warehouse .....	24
6.2 Carga .....	25
6.2.1 Criar um pipeline de dados .....	25
6.3 ETL.....	29
6.3.1 Criação do processo de ETL .....	29
6.3.2 Consultas Visuais para fluxo ETL .....	30
6.3.3 Consultas Visuais criadas.....	31
6.3.4 Criação de Views.....	35
6.3.5 Gravação dos dados em um novo banco de dados .....	39
7. Análise.....	43
7.1 Problemas .....	44
7.2 Resolução.....	47
8. Análise dos objetivos estabelecidos.....	50

# 1. SOLUÇÃO ESCOLHIDA

Durante os estudos referentes às soluções para ETL em ambientes de nuvem, surgiu o interesse de analisar o recente lançamento do “Microsoft Fabric”.

A proposta inicial do Microsoft Fabric visa fornecer uma abordagem integrada e robusta para a gestão de dados, possibilitando a flexibilidade de explorar dados em seu formato bruto ou estruturado e capacitar as empresas a tomar decisões estratégicas e informadas. O Fabric reúne experiências como Engenharia de Dados, Data Factory, Ciência de Dados, Data Warehouse, Análise em Tempo Real e Power BI em uma base SaaS compartilhada.

## A solução propõe:

- **Unificação de Dados:**  
Integrar e unificar dados de várias fontes e formatos para criar um repositório de dados unificado e acessível;
- **Flexibilidade de Análise:**  
Fornecer flexibilidade na análise de dados, permitindo consultas em dados brutos e estruturados;
- **Eficiência Operacional:**  
Melhorar a eficiência operacional por meio de automação, gerenciamento simplificado e recursos de governança de dados;
- **Inteligência de Negócios:**  
Capacitar as organizações a obter insights valiosos e tomar decisões informadas por meio de análises avançadas e recursos de *Business Intelligence*;
- **Escalabilidade e Segurança:**  
Oferecer escalabilidade para lidar com grandes volumes de dados e garantir alta segurança e conformidade dos dados.

Sendo assim, uma análise aprofundada de uma proposta inovadora e recente justifica a adoção do Microsoft Fabric para avaliar como essa solução atende às demandas de gestão de dados em um ambiente empresarial em constante transformação.



Figura 1 - Componetes do Microsoft Fabric



## 2. OBJETIVO

A empresa Wide World Imports enfrentou alguns desafios, nos quais os gestores vieram a perceber que a empresa estava fora do cenário atual, que busca um acesso fácil e em escalabilidade para a gestão dos dados.

Dessa forma, tomou a decisão de implementar o processo de alocação dos dados em nuvem, a fim de focar em seus objetivos que visam:

- melhorar a eficiência operacional;
- impulsionar a inovação e proporcionar vantagens competitivas;
- redução de custos
- colaboração global;
- segurança, escalabilidade e conformidade nos dados;
- Integração e unificação dos dados de várias fontes e formatos para criar um repositório de dados unificado e acessível;
- Fornecer flexibilidade na análise de dados, permitindo consultas em dados brutos e estruturados;
- Criar processos de Governança de Dados
- Embasar os Negócios em informações de Inteligência / insights

Inicialmente, os gestores propuseram uma implantação gradual, desde a estruturação dos dados.

Com o propósito de estimular reflexões analíticas, foram apresentadas as seguintes questões:

1. Dos quatro anos disponíveis para análise, qual foi o melhor ano, em termos de faturamento?
2. Do catálogo de produtos, há algum produto que não possuí vendas?
3. No último ano, como estão as vendas? Há variação positiva?
4. Qual é o cliente que mais faturou?
5. No último ano, quantos produtos representam 80% do meu faturamento?
6. Em quais produtos / grupos possuo maior margem bruta?
7. Qual é a projeção para os próximos meses?

### 3. ESCOLHA DA BASE DE DADOS

Escolheu-se a base de dados Wide World Importers sample, cujo arquivo necessita ser o bacpac, já que a restauração foi feita na Azure.

O link abaixo, é para acesso ao ambiente de download:

<https://github.com/Microsoft/sql-server-samples/releases/tag/wide-world-importers-v1.0>

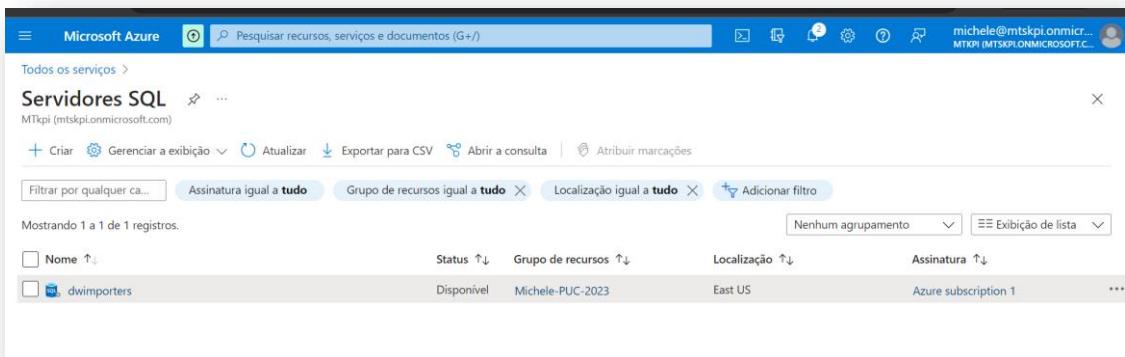


# 4. COLETA

Uma vez que o download foi realizado, optou-se por hospedar os dados na Azure, para isso, seguiu-se as seguintes etapas:

## 4.1 RESTAURAÇÃO DO BANCO DE DADOS

- O processo de restauração, inicia-se com o download da versão “bacpac”;
- Antes de restaurar, é necessário criar o servidor de banco de dados, no ambiente Azure



Nome	Status	Grupo de recursos	Localização	Assinatura
dwimporters	Disponível	Michele-PUC-2023	East US	Azure subscription 1

Figura 2 - Servidor de Bancos SQL no Azure

Para facilitar a gestão, criou-se o grupo de Recursos, de forma que todos os recursos, ficarão aloados nesse grupo

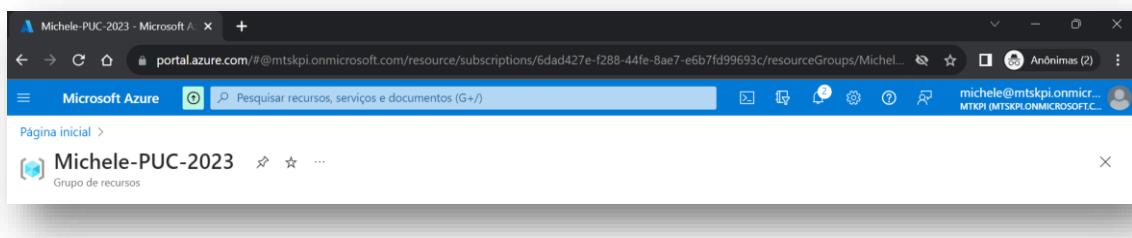


Figura 3 - Grupo de Recursos

- A restauração tem que ocorrer pelo SQL Server Management Studio



Figura 4 - SSM

- d) No print seguinte, a restauração já havia ocorrido, mas para restaurar é necessário conectar no servidor da Azure e escolher a opção “Importar Aplicativo da Camada de Dados”, apontando para o arquivo bacpac;

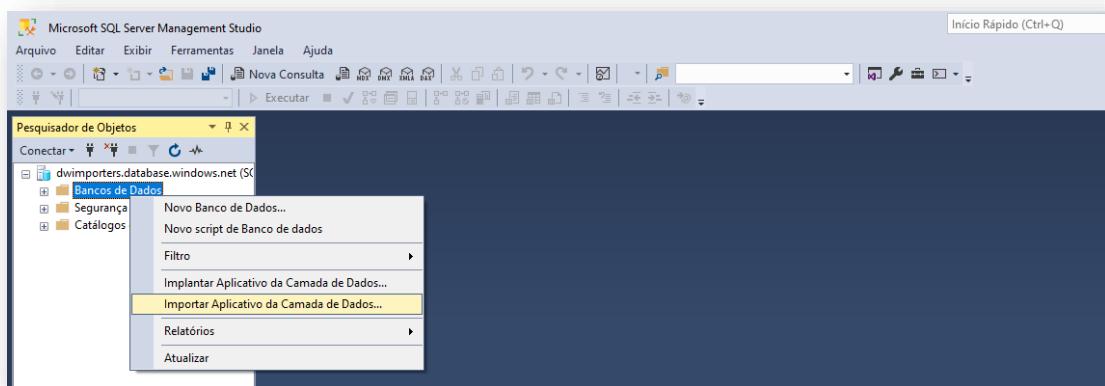


Figura 5 - Opção para restauração de bancos no Azure

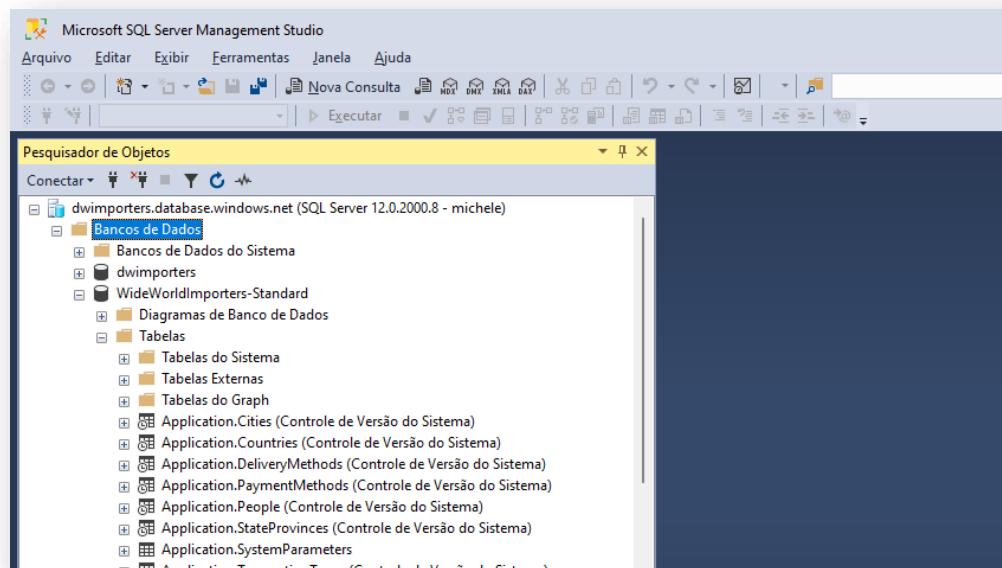


Figura 6 - Amostra do banco restaurado

The screenshot shows the Microsoft Azure portal interface for the 'WideWorldImporters-Standard' database. The left sidebar has a navigation menu with options: Visão geral, Log de atividade, Marcações, Diagnosticar e resolver problemas, Editor de consultas (visualização), Configurações, Computação + armazenamento, Cadeias de conexão, Propriedades, and Bloqueios. Under 'Editor de consultas (visualização)', there is a note: 'Mostrando o explorador de objetos limitados aqui. Para obter a capacidade completa, clique aqui para abrir o Azure Data Studio.' The main content area shows the database structure with a tree view under 'WideWorldImporters-Standard (michele)'. The 'Tabelas' node is expanded, listing tables such as Application.Cities, Application.Cities\_Archive, Application.Countries, Application.Countries\_Archive, Application.DeliveryMethods, Application.DeliveryMethods\_Archive, Application.PaymentMethods, Application.PaymentMethods\_Archive, and Application.People. On the right, a query editor window titled 'Consulta 1' is open, showing a single result row with the number '1'. The 'Resultados' tab is selected, and there is a search bar at the bottom.

Figura 7 - Banco WideWorldImporters-Standard – Azure



# 5. MODELAGEM E CATÁLOGO DE DADOS

- Construção de um Modelo de Dados:

Foi desenvolvido um modelo de dados, com a opção Esquema Estrela, sendo que a construção (via ETL) iniciou-se no Warehouse do Microsoft Fabric, sendo hospedado, por fim, em um banco SQL Server, na Azure.

- Documentação da Linhagem dos Dados:

Foi elaborada uma documentação que descreve a linhagem dos dados, com informações desde a origem dos dados.

O banco de dados, original, foi obtido através do arquivo bacpac, conforme descrito no capítulo sobre escolha da base de dados.

- Criação de um Catálogo de Dados:

Um Catálogo de Dados foi criado como parte do processo. Este catálogo continha, no mínimo, uma descrição detalhada dos dados e seus domínios.

Essas atividades foram essenciais para garantir a qualidade e a rastreabilidade dos dados no contexto do projeto. O modelo de dados escolhido (Esquema Estrela) e o Catálogo de Dados forneceram estruturas sólidas para armazenar e gerenciar informações, enquanto a documentação da linhagem dos dados ajudou a manter a transparência e a confiabilidade dos dados utilizados no projeto.

## 5.1 MODELAGEM

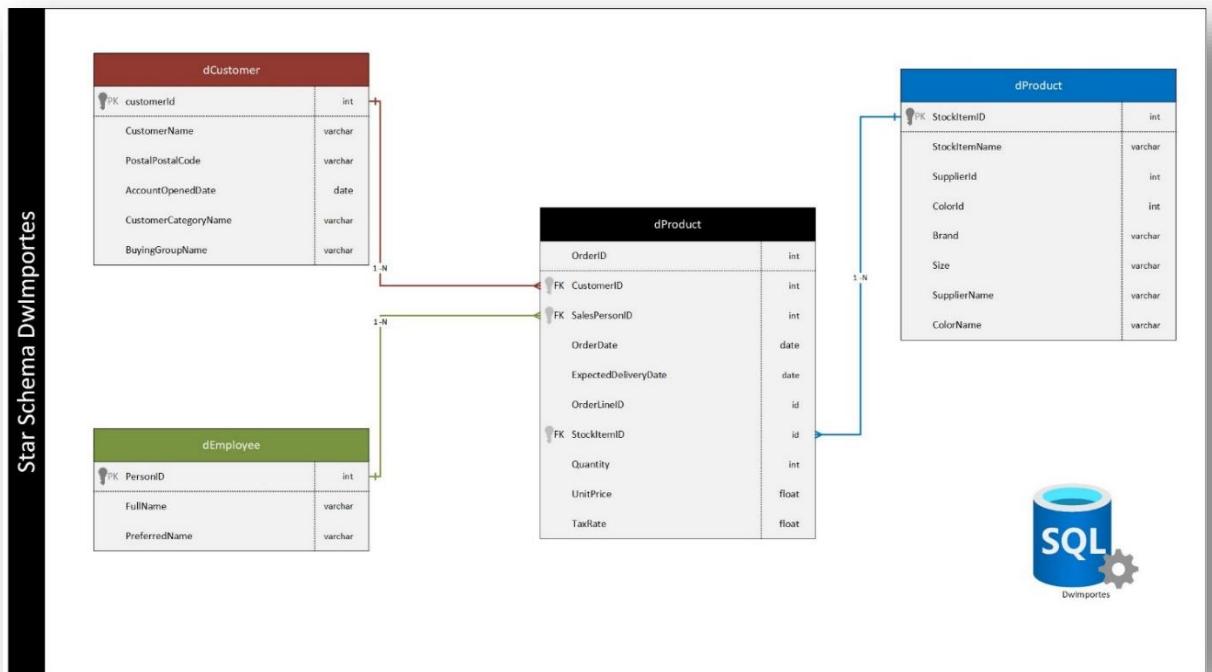


Figura 8 – Modelo Star Schema proposto

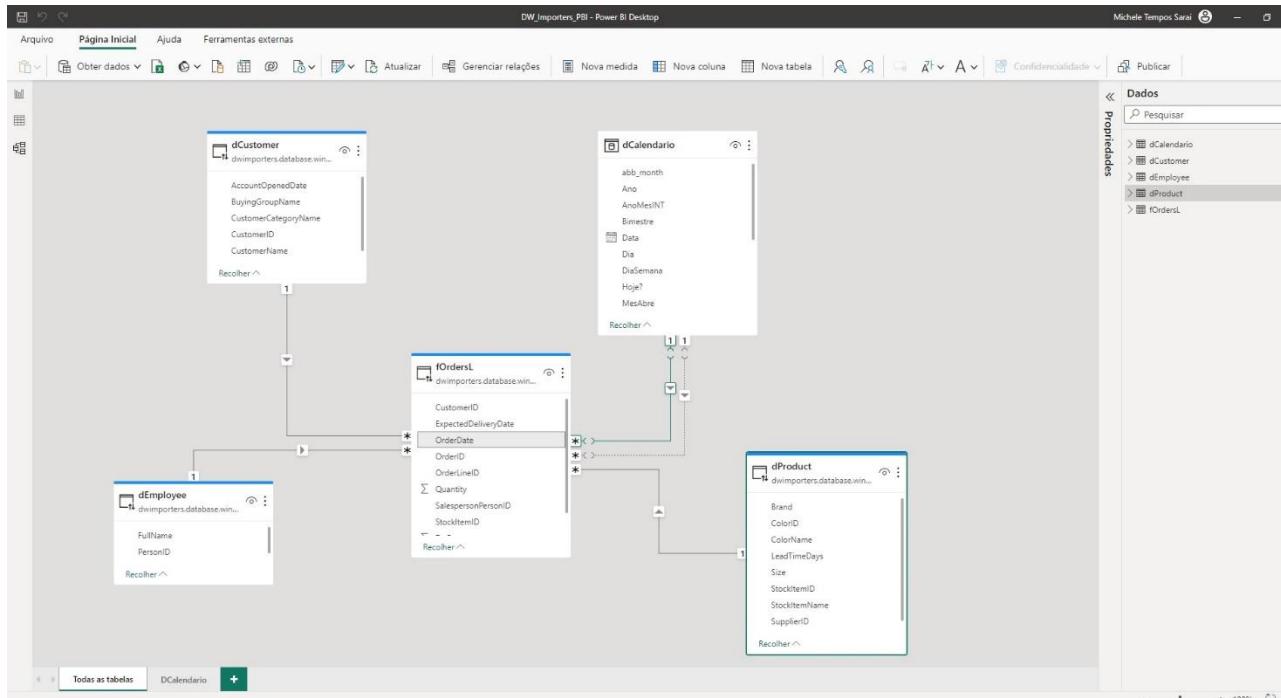


Figura 9 - Modelo Star Schema pronto

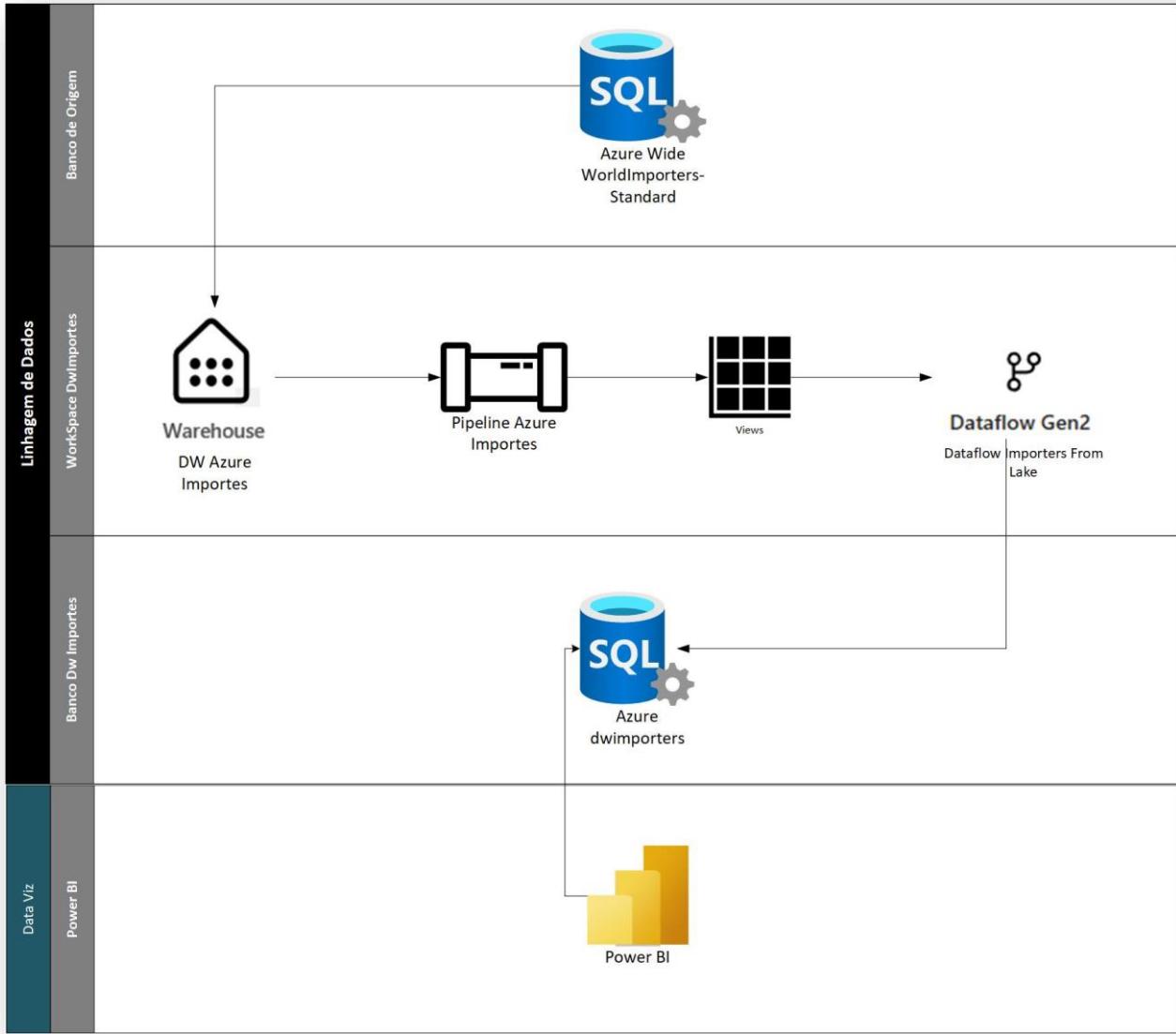


Figura 10 – Linhagem de Dados – ETL

## 5.2 CATALOGANDO OS DADOS

Mantendo-se na opção de trabalhar os dados na nuvem, optou-se pela catalogação dos dados através do Microsoft Purview.

Conforme a própria Microsoft define: “O Microsoft Purview fornece uma solução de governança de dados unificada para ajudar a gerenciar e controlar seus dados locais, multinuvem e SaaS (software como serviço).”

Assim, é possível criar o mapeamento dos dados, visualizar a hierarquia, realizar a classificação de dados, inclusive apontando os confidenciais, configurar a descoberta de automatizada de dados, indicar responsáveis pelos dados, criar glossários. Enfim, realizar toda configuração de

metadados. Assim, é possível facilitar a pesquisa para localizar os dados, filtrando os resultados da pesquisa por termos de negócios, classificações e contatos.



Figura 11 – Ferramentas de governança disponíveis no Microsoft Purview

A screenshot of the Microsoft Azure portal showing the 'Contas do Microsoft Purview' (Microsoft Purview Accounts) page. The page displays a table with one row of data:

Nome	Tipo	Grupo de recursos	Localização	Assinatura	Status
Pw-Michele	Conta do Microsoft Purview	Michele-PUC-2023	Brazil South	Azure subscription 1	Bem-sucedido

Figura 12 – Conta do Microsoft Purview criada/configurada

### 5.2.1 Mapa Unificado dos Dados

Para identificação das fontes de dados, é possível criar o Mapa de Dados

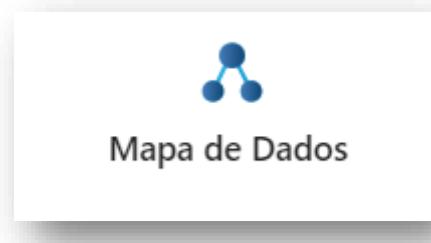


Figura 13 – Mapa de Dados – Microsoft Purview

Com essa opção, é possível visualizar, de forma unificada, todas as origens de dados que forem mapeadas. Caso haja mais fontes, elas aparecerão uma ao lado da outra

Microsoft Purview PREVIEW

Página Inicial

Mapa de Dados

Fontes de dados

Coleções

Monitoramento

Glossário de negócios

Metamodel

Gerenciamento de origem

Gerenciamento de anota...

Aplicativos

Aplicativos relacionados

Catálogo de Dados

Insights do Patrimônio de Da...

Exibição do mapa

Exibição da tabela

Filtrar por palavra-chave

Pw-Michele  
The root collection.

AzureSqlDatabase-Purv...  
Banco de Dados SQL do Azure

Ver os detalhes

Figura 14 – Mapeamento visual

**AzureSqlDatabase-Purview**  
Banco de Dados SQL do Azure

**Visão Geral**

ID da fonte de dados: dwimporters.database.windows.net

Exames Ativos descobertos Ativos confidenciais

Nome do exame	Status da últ... Concluído	Conjunto de r... AzureSqlDatab...	Hora do último e... 10/09/2023, 19:32	Extração de linhagem Desativado
ScanMichele	<span>Ver todos os exames aplicados</span>			

Verificações com erros recentes

Nome do exame	Status	Sugestões
Nenhuma execução de exame recente com falha		

→ Ver todas as execuções de exame com falha

Registrado em 10/09/2023, 17:56:29

Caminho da coleção Pw-Michele

Hierarquia de origem

- Azure subscription 1
- Assinatura
- Michele-PUC-2023
- Grupo de recursos

Figura 15 – Detalhes do Mapeamento

**Sources**

Showing 5 collections, 1134 sources

Collection	Source Type	View details
NorthAmericaDataCenter	SQL Server	View details
EuropeDataCenter	SAP S/4HANA (Preview)	View details
AzureAndBINorthAmerica	Azure Data Lake Storage Gen2	View details
AmazonNorthAmerica	Azure Blob Storage	View details
AzureEurope	Amazon S3	View details

Map view

Figura 16 – Exemplo disponibilizado pela Microsoft

### 5.2.1 Catálogo de Dados

Para descrever os ativos de dados, utiliza-se Catálogo de Dados



Figura 17 - Catálogo de Dados - Microsoft Purview

A screenshot of the Microsoft Purview Catalogue interface. The left sidebar shows navigation options like 'Visão Geral', 'Navegar', 'Compartilhamento de d...', and 'Aplicativos'. The main area displays a 3D visualization of data assets, a search bar, and sections for 'Explorar seus dados' (with a 'Microsoft Azure' card) and 'Acesso rápido' (listing recent items like 'dEmployee', 'dCustomer', 'dOrdersL', and 'dProduct').

Figura 18 – Catálogo de Dados do Microsoft Purview

Na opção de catálogo de dados, é possível descrever os ativos identificados. Por exemplo, a tabela `dCustomer` é um ativo, no qual há uma descrição breve, apontamento de classificações para dados sensíveis, pessoais (isto é, dados normalmente protegidos por leis específicas, como a LGPD) e ainda é possível descrever campo a campo do ativo, em uma linguagem natural e de fácil interpretação (em schemas)

The screenshot shows the Microsoft Purview Catalog interface. On the left, there's a sidebar with navigation links like 'Página Inicial', 'Catálogo de Dados' (selected), 'Visão Geral', 'Navegar', 'Compartilhamento de d...', 'Aplicativos', 'Aplicativos relacionados', 'Mapa de Dados', and 'Insights do Patrimônio de Da...'. The main content area is titled 'Catálogo de Dados > dCustomer'. It shows a table structure with columns: Nome da coluna, Classificações, Rótulo de co..., Termos do gl..., Tipo de dados, and Descrição da coluna. The table contains six items: CustomerID (bigint, primary key, cliente código), CustomerName (nvarchar, cliente nome), PostalPostalCode (nvarchar, principal endereço CEP), AccountOpenedDate (date, criação data), CustomerCategoryName (nvarchar, categoria cliente), and BuyingGroupName (nvarchar, organização cliente). There are also classification dropdowns for CustomerName ('All Full Names', 'U.S. Zip Codes') and PostalPostalCode ('All Full Names'). At the top right, there are buttons for 'Abrir no Power BI Desktop' and other actions.

Figura 19 - Tabela dCustomer - em catálogo de dados

O Microsoft Purview disponibiliza várias classificações, de diversos países. Ainda sim, é possível criar classificações próprias.

The screenshot shows the Microsoft Purview 'Classificações' (Classifications) page. The left sidebar includes 'Coleções', 'Monitoramento', 'Glossário de negócios', 'Metamodel', 'Ativos de negócios (versão p...)', 'Tipos de ativos (versão prévia)', 'Atributos gerenciados', 'Gerenciamento de orig...', 'Verificar conjuntos de regras', 'Regras de padrão', 'Runtimes de integração', 'Credenciais', 'Conexões de linhagem', 'Identidades gerenciadas', 'Conexões', 'Gerenciamento de ano...', 'Classificações' (selected), 'Regras de classificação', 'Aplicativos', 'Aplicativos relacionados', 'Catálogo de Dados', and 'Insights do Patrimônio de Da...'. The main content area is titled 'Classificações' and shows a table with columns: Nome de exibição, Nome formal, and Descrição. It lists various classifications from different countries, such as 'ABA Routing Number', 'All Full Names', 'All Physical Addresses', 'Argentina National Identity (DNI) Number', 'Australia Bank Account Number', 'Australia Business Number', 'Australia Company Number', 'Australia Driver's License Number', 'Australia Medical Account Number', 'Australia Passport Number', 'Australia Tax File Number', 'Austria Driver's License Number', 'Austria Identity Card', 'Austria Passport Number', 'Austria Social Security Number', and 'Austria Tax Identification Number'. The table also includes a note: 'Essas são as classificações fornecidas pelo sistema.' (These are the classifications provided by the system.)

Figura 20 – Classificações para dados Sensíveis/Pessoais

No mesmo exemplo da dCustomer, veja como é o preenchimento da descrição do ativo e uma demonstração de possíveis valores para categorias.

Figura 21 – Descrição do ativo + possíveis categorias

No print seguinte, é demonstrado dois campos onde há um apontamento de classificação de dados sensíveis, `CustomerName` e `PostalCode`. É possível fazer mais de uma classificação por campo (esquema)

Figura 22 – Classificação de dados sensíveis

## Veja outros exemplos catalogados

The screenshot shows the Microsoft Purview Catalog interface for the **dCustomer** database. The left sidebar includes links for Página Inicial, Catálogo de Dados (selected), Visão Geral, Navegar, Compartilhamento de dados, Aplicativos, and Aplicativos relacionados. The main content area displays the **Catálogo de Dados > dCustomer**. It shows a tree view with **dwimporters.database.windows.net** (SQL Server do Azure) containing **dwimporters** (Banco de Dados SQL do Azure) and **dbo** (Esquema do Azure SQL). A message indicates there are no related items. Below the tree view is a search bar labeled **Pesquisar**.

Figura 23 – Demais tabelas relacionadas

The screenshot shows the Microsoft Purview Catalog interface for the **dEmployee** database. The left sidebar includes links for Página Inicial, Catálogo de Dados (selected), Visão Geral, Navegar, Compartilhamento de dados, Aplicativos, and Aplicativos relacionados. The main content area displays the **Catálogo de Dados > dEmployee**. It shows a tree view with **dwimporters.database.windows.net** (SQL Server do Azure) containing **dwimporters** (Banco de Dados SQL do Azure) and **dbo** (Esquema do Azure SQL). A message indicates there are 3 of 3 items. Below the tree view is a search bar labeled **Filtrar por nome** and a table showing column details:

Nome da coluna	Classificações	Rótulo de co...	Termos do gl...	Tipo de dados	Descrição da coluna
PersonID				bigint	Código do funcionário
FullName	All Full Names			nvarchar	Nome completo do funcionário
PreferredName	All Full Names			nvarchar	Nome de preferência, apelido

Figura 24 – dEmployee

Microsoft Purview

Catálogo de Dados > dProduct

Visualizar no SQL do Azure + Adicionar Marca

Visão Geral Selecione para edição em massa Solicitar acesso Atualizar Excluir Editar Colunas

Filtrar por nome

Propriedades Mostrando 9 de 9 itens

Esquema Nome da coluna Classificações Rótulo de coluna Termos de g... Tipo de dados Descrição da coluna

Linha	StockItemID			bigint	Código do produto
Contatos	StockItemName			nvarchar	Nome/Descrição do produto
Relacionado	SupplierID			bigint	Código do Fornecedor, chave estrangeira com a tabela Supplier
	ColorID			bigint	Código da cor
	Brand			nvarchar	Marca do produto
	Size			nvarchar	Tamanho do produto - alguns produtos não possuem tamanhos
	LeadTimeDays			bigint	Tempo, em dias, para a entrega do produto pelo fornecedor Mínimo: 1 Máximo: 30
	SupplierName			nvarchar	Nome do Fornecedor
	ColorName			nvarchar	Nome da cor

Atualizado em 12 de setembro de 2023 às 19:59 por Michele Tempis Sarai

Abrir no Power BI Desktop

Figura 25 – dProduct

Microsoft Purview

Catálogo de Dados > fOrdersL

Visualizar no SQL do Azure + Adicionar Marca

Visão Geral Selecione para edição em massa Solicitar acesso Atualizar Excluir Editar Colunas

Filtrar por nome

Propriedades Mostrando 10 de 10 itens

Esquema Nome da coluna Classificações Rótulo de coluna Termos de g... Tipo de dados Descrição da coluna

Linha	OrderID			bigint	Número do pedido de venda
Contatos	CustomerID			bigint	Chave estrangeira referente ao código do Cliente (dCustomers)
Relacionado	SalespersonPersonID			bigint	Chave estrangeira referente ao código do Funcionário (dEmployee)
	OrderDate			date	Data de realização dos pedidos de venda
	ExpectedDeliveryDate			date	Data de prazo de expectativa de envio dos produtos do pedido
	OrderLineID			bigint	Número do item do pedido
	StockItemID			bigint	Chave estrangeira correspondente ao código do Produto (dProduct)
	Quantity			bigint	Quantidade de produtos do item do pedido
	UnitPrice			float	Preço unitário de venda, do item do pedido
	TaxRate			float	Taxas/Impostos

Atualizado em 12 de setembro de 2023 às 20:01 por Michele Tempis Sarai

Abrir no Power BI Desktop

Figura 26 – fOrdersL – tabela fato

# 6. CARGA

Nesta etapa, foi realizada a carga de dados para o Data Warehouse do Microsoft Fabric. A seguir, estarão descritos os processos para a carga e transformação dos dados (ETL)

## 6.1 WORKSPACE

É necessário criar um WorkSpace para iniciar o processo de carga. Para isso, deve-se acessar o ambiente do Power BI e criar o WorkSpace

The screenshot shows the Power BI workspace creation interface. On the left, there's a sidebar with options like 'Pagina inicial', 'Criar', 'Pesquisar', 'Hábitos de dados do Power BI', and 'Workspaces'. Under 'Workspaces', it says 'DW\_Importers' and has a 'Novo workspace' button. The main area is titled 'Power BI | DW\_Importers' and shows a table of workspaces. One row for 'DW\_Importers' is highlighted with a red arrow pointing to it. The table columns are 'Tipo', 'Proprietário', 'Atualizado', 'Próxima atualização', and 'Endoso'. The 'DW\_Importers' row contains: 'Fluxo de dados Gen2', 'Michele Tempos Sarai', '07/09/23, 09:48:38', 'N/D', and an empty endoso field.

Figura 27 – Criação do WorkSpace (DW\_Importers)

### 6.1.1 Visualizando as opções do Fabric

Após acessar o WorkSpace, podemos clicar na opção Novo >> Mostrar tudo. Assim, é possível visualizar todas as opções que o Fabric disponibiliza, inclusive o Data Warehouse

Power BI DW\_Importers

**DW\_Importers**

+ Novo Carregar Criar aplicativo Gerenciar acesso Configurações de...

- Relatório
- Relatório paginado
- Scorecard
- Painel
- Conjunto de Dados
- Fluxo de dados
- Datamart (Visualização)
- Conjunto de dados de streaming
- Fluxo de dados de streaming

Mostrar tudo

Tipo	Proprietário
Fluxo de dados Gen2	Michele Tempos Sa...
Fluxo de dados Gen2	Michele Tempos Sa...
Conjunto de dados (p: DW_Importers)	
Ponto de extremidade DW_Importers	
Lakehouse	Michele Tempos Sa...

Figura 28 – Mostrar tudo – no WorkSpace

Power BI DW\_Importers

Novo

Workshop atual: DW\_Importers

Data Engineering

Crie um lakehouse e operacionalize seu fluxo de trabalho para criar, transformar e compartilhar seu patrimônio de dados.

Lakehouse (Visualização) Ambiente de Data para limpeza, consulta, relatório e compartilhamento.

Bloco de Anotações (Visualização) Explore dados e crie soluções de aprendizado de máquina com os aplicativos do Apache Spark.

Definição de Trabalho do Spark Defina, program e gerencie seu trabalho de Apache Spark para processamento de Big Data.

Data Factory

Capacite sua organização a obter o valor dos dados mais rápido do que nunca.

Fluxo de dados Gen2 (Versão ...) Preparar, Importar e transformar dados.

Pipeline de dados (Visualização) Ingerir dados em escala e agendear fluxos de trabalho de dados.

Data Science

Use o aprendizado de máquina para detectar tendências, identificar valores atípicos e prever valores a partir dos seus dados. [Saiba mais](#).

Modelo (Visualização) Use modelos de machine learning para prever resultados e detectar anomalias nos dados.

Experimento (Visualização) Crie execução e acompanhe o desenvolvimento de vários modelos para validar hipóteses.

Bloco de Anotações (Visualização) Explore dados e crie soluções de aprendizado de máquina com os aplicativos do Apache Spark.

Data Warehouse

Forneca insights estratégicos de várias fontes para todo o seu negócio. [Saiba mais](#).

Warehouse (Visualização) Forneca insights estratégicos de várias fontes para todo o seu negócio.

Power BI

Use ótimos para encontrar insights, acompanhar o progresso e tomar decisões mais rapidamente. [Saiba mais](#).

Relatório Crie uma apresentação interativa dos seus dados.

Relatório paginado Edite dados tabulares em um relatório fácil de importar e compartilhar.

Scorecard Defina, acompanhe e compartilhe as principais métricas da sua organização.

Painel Crie uma história de dados de uma única página.

Fluxo de dados Preparar, Importar e transformar dados.

Figura 29 – Opção Data Warehouse

Outra maneira, é através dos botões de acesso fácil, no canto inferior esquerdo

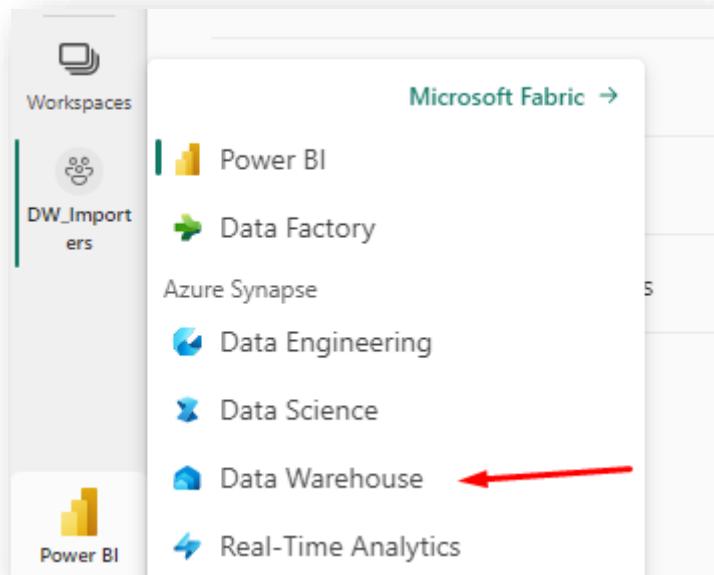


Figura 30 – Botões de acesso fácil

### 6.1.2 O Data Warehouse

A criação do Data Warehouse é bem simples, veja o preenchimento abaixo

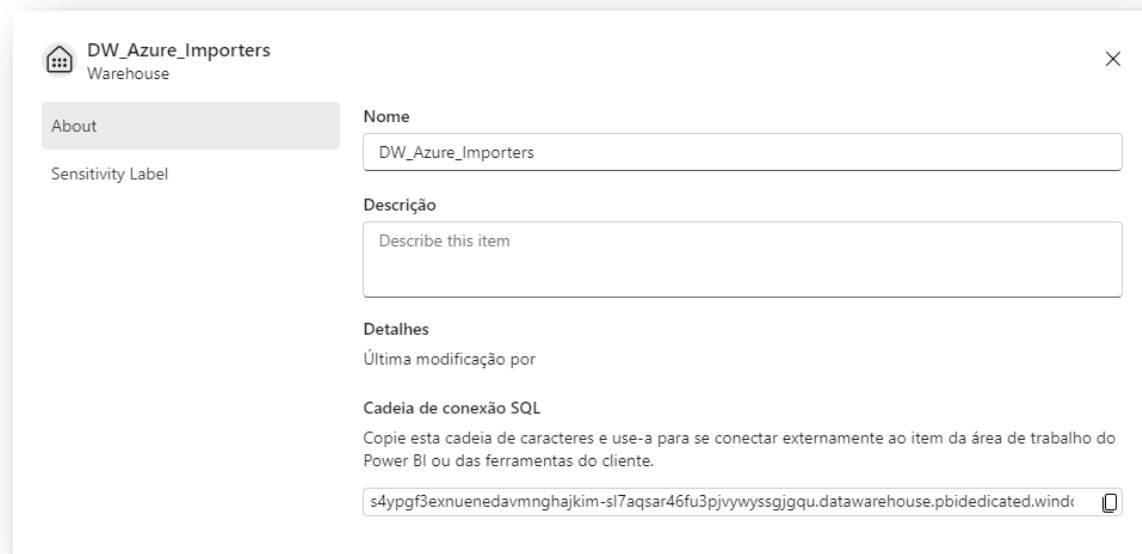


Figura 31 – Criação do Data Warehouse

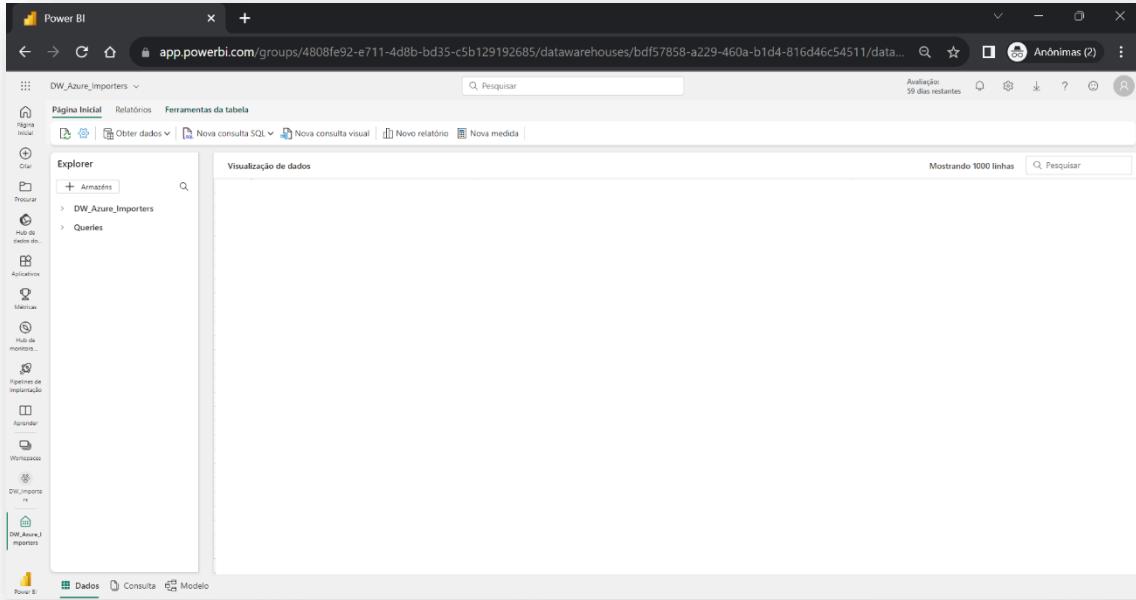


Figura 32 – Data Warehouse criado

## 6.2 CARGA

### 6.2.1 Criar um pipeline de dados

Umas das formas de ingerir dados para o Data Warehouse, é através do Pipeline de Dados

Dentro do DW, escolha a opção “Obter Dados” → “Novo pipeline de dados”

Escolha a fonte de dados e realize o preenchimento das informações para conexão

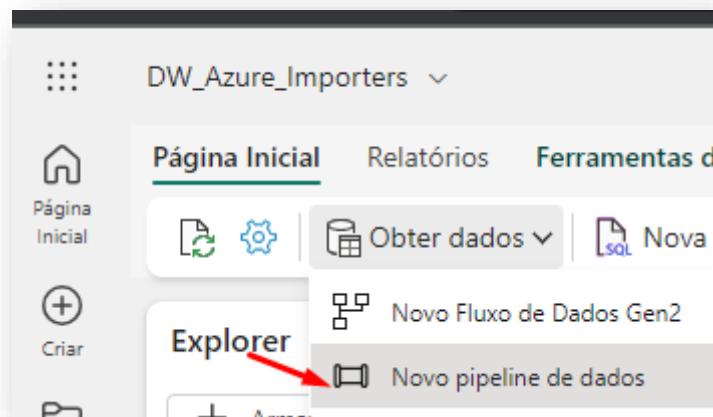


Figura 33 – Pipeline de Dados

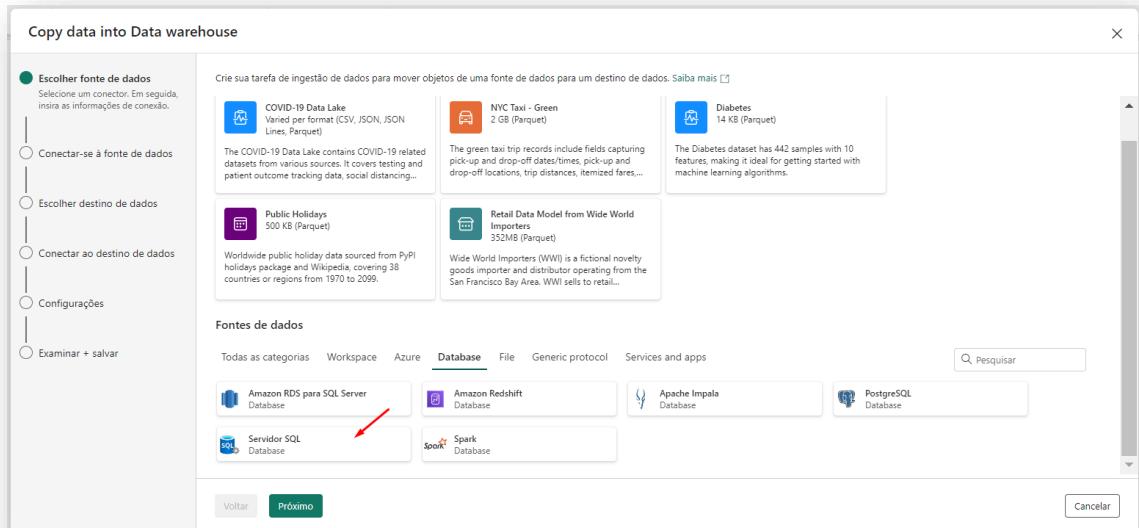


Figura 34 – Fonte de dados

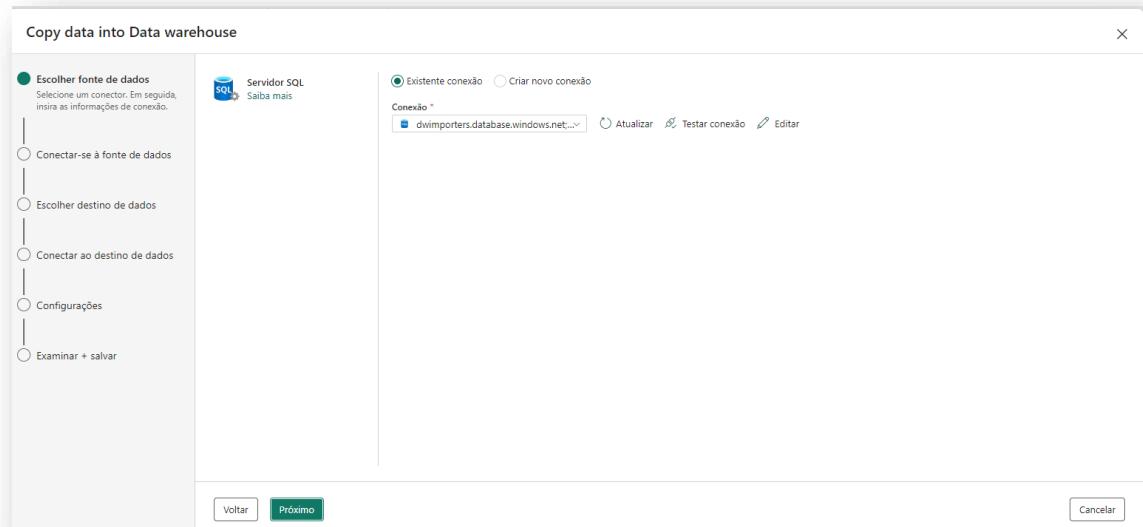


Figura 35 – Conexão com banco SQL Server Azure

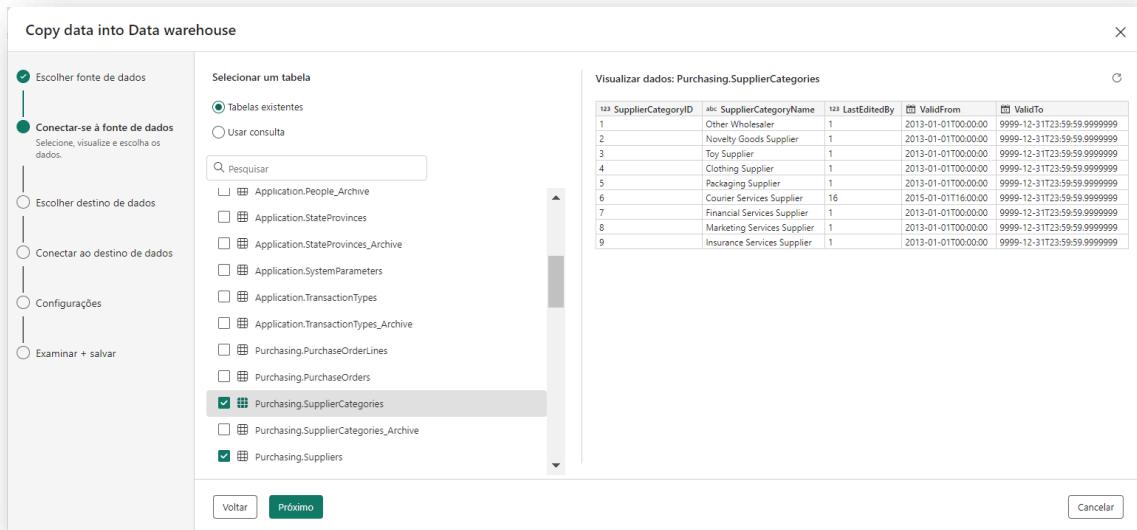


Figura 36 – Selecionando as fontes de dados

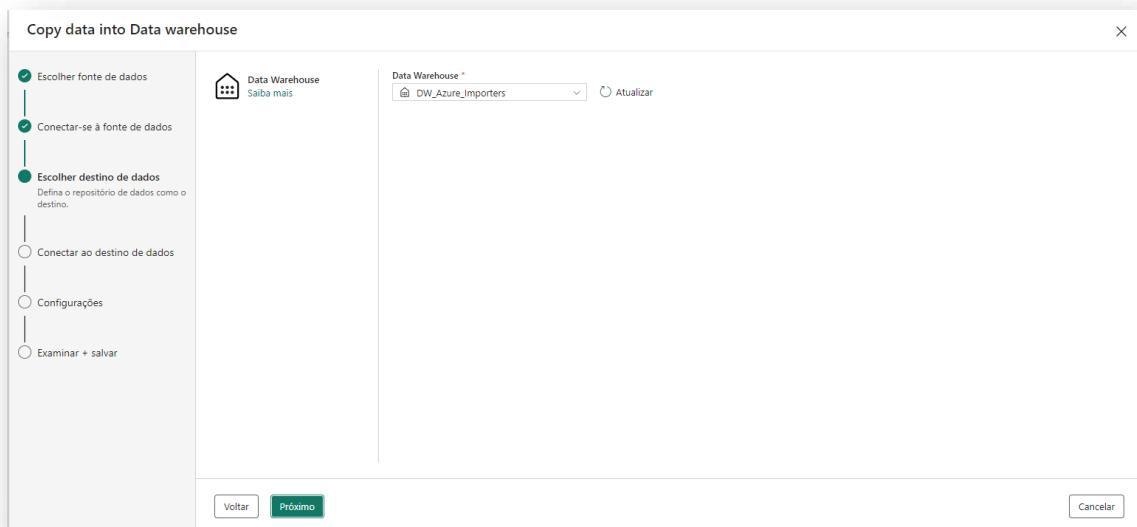


Figura 37 – Cópia das tabelas para o DW

Nesta etapa, já é possível definir quais colunas são necessárias na importação

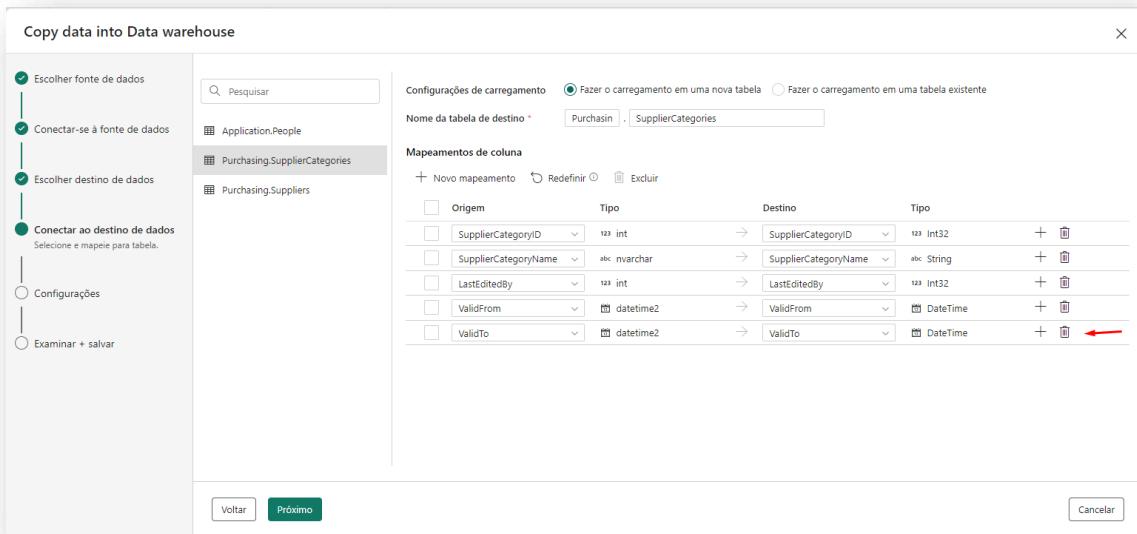


Figura 38 – Opção para exclusão de colunas não necessárias

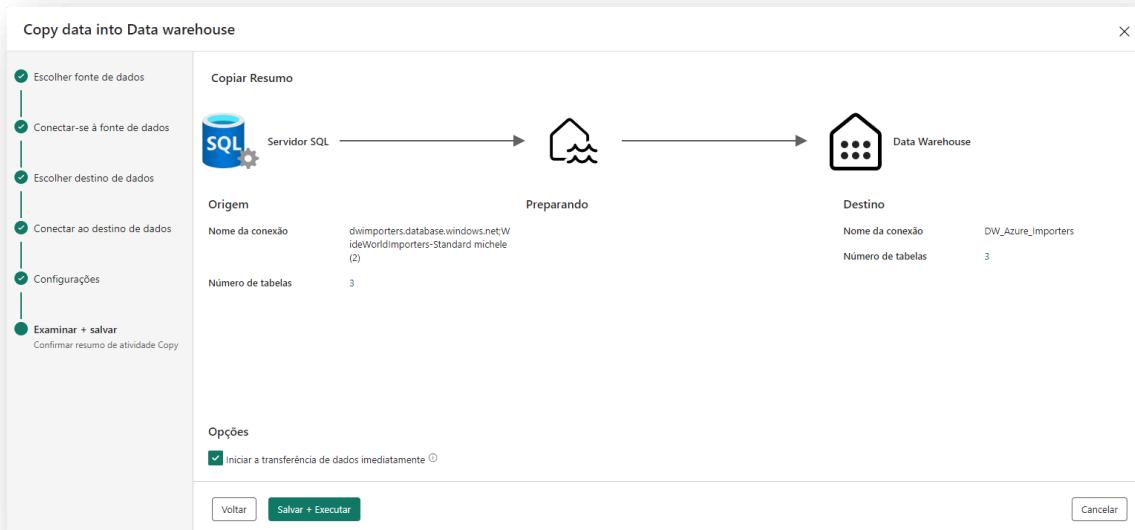


Figura 39 – Etapa final de verificação

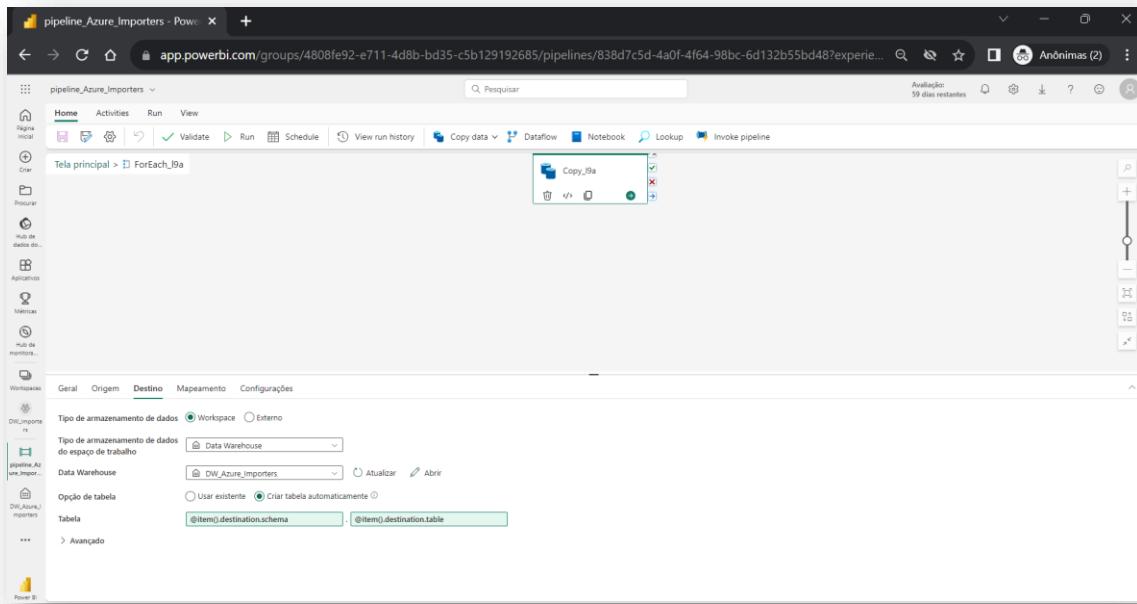


Figura 40 – Pipeline criado

## 6 . 3 E T L

### 6 . 3 . 1 C r i a ç ã o d o p r o c e s s o d e E T L

Uma vez que a carga da origem foi destinada para o Data Warehouse, é possível criar processos de ETL, nos quais, as transformações dos dados ocorrerão.

Ao todo, foram transformadas 9 tabelas, que deram origem a 4 tabelas, no formato Star Schema, note que as tabelas ficam disponíveis do lado esquerdo

	CustomerID	CustomerName	CustomerCategoryID	BuyingGroupID	PostalPostalCode
1	1	Talstein Toys (Head Office)	3	1	90410
2	2	Talstein Toys (Sylvania, MT)	3	1	90216
3	3	Talstein Toys (Peoples Valley, AZ)	3	1	90205
4	4	Talstein Toys (Medicine Lodge, KS)	3	1	90152
5	5	Talstein Toys (Georgetown, NV)	3	1	90261
6	6	Talstein Toys (Leisure, ND)	3	1	90298
7	7	Talstein Toys (Franklinville, TN)	3	1	90761
8	8	Talstein Toys (Bonne Mac, CO)	3	1	90404
9	9	Talstein Toys (Netcong, NJ)	3	1	90129
10	10	Talstein Toys (Wimbledon, ND)	3	1	90061
11	11	Talstein Toys (Deauville, PA)	3	1	90185
12	12	Talstein Toys (Biscay, MN)	3	1	90054
13	13	Talstein Toys (Stonefort, IL)	3	1	90685
14	14	Talstein Toys (Long Meadow, MD)	3	1	90633
15	15	Talstein Toys (Baton, TN)	3	1	90631
16	16	Talstein Toys (Coney Island, MO)	3	1	90467
17	17	Talstein Toys (East Fultonham, OH)	3	1	90416
18	18	Talstein Toys (Goffstown, NH)	3	1	90321
19	19	Talstein Toys (Lemetta, AK)	3	1	90303
20	20	Talstein Toys (College Place, WA)	3	1	90050
21	21	Talstein Toys (Treckover, RI)	3	1	90790
22	22	Talstein Toys (Ward Ridge, FL)	3	1	90764
23	23	Talstein Toys (Ikaian, AK)	3	1	90019
24	24	Talstein Toys (Dundarrach, NC)	3	1	90758
25	25	Talstein Toys (Arenal, CA)	3	1	90352
26	26	Talstein Toys (Herrick, MA)	3	1	90331

Figura 41 – Tabelas disponíveis para o processo de ETL

### 6.3.2 Consultas Visuais para fluxo ETL

No ambiente do DW, existem as consultas visuais, que possibilitam montar todas as etapas do ETL de forma visual.

Os componentes podem ser adicionados, na sequência, criando um fluxo. Caso necessite voltar e alterar um componente, é só clicar duas vezes sobre ele. Abaixo, um exemplo de edição em um componente de “Join”

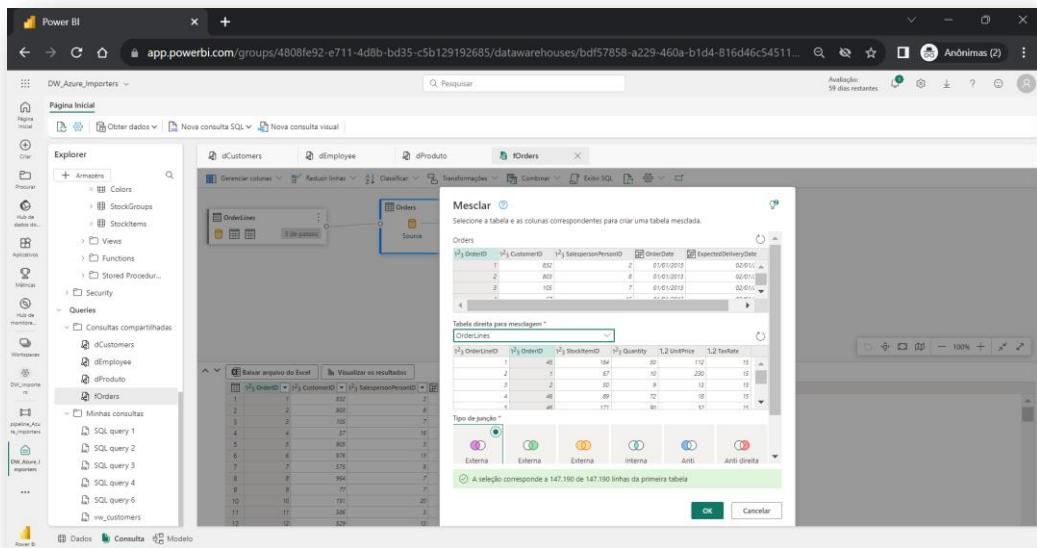


Figura 42 – Componente “Join”

Um outro exemplo, é a escolha de colunas que serão mantidas na exibição final. Após a inserção de um processo de “Join”, é natural que algumas colunas apareçam em duplicidade. No exemplo seguinte, mostra-se a escolha das colunas que permanecerão

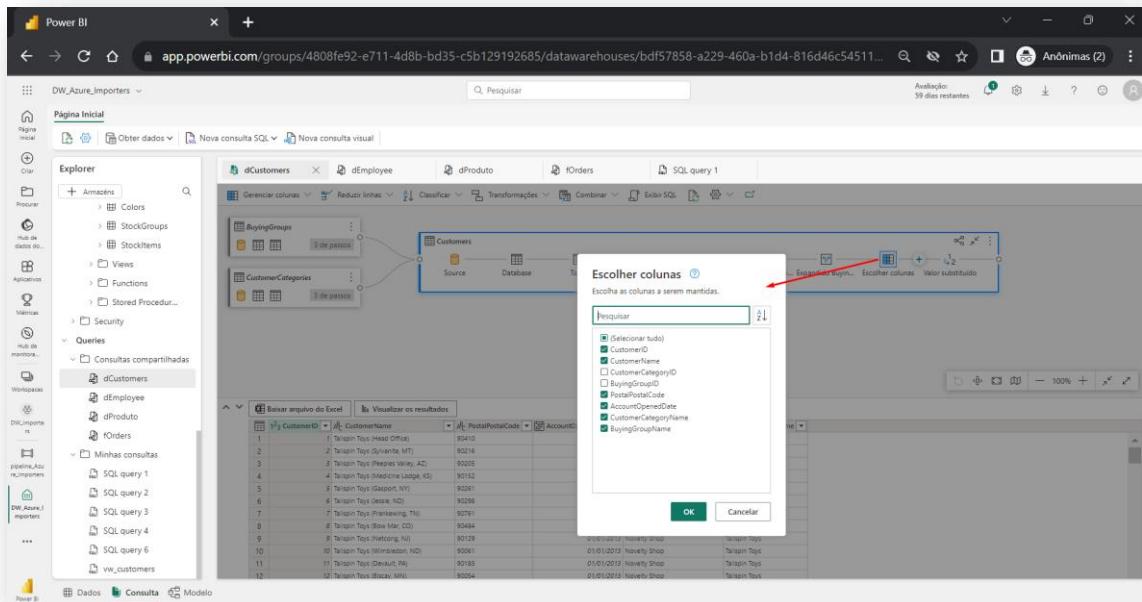


Figura 43 – Componente “Escolher Colunas”

### 6.3.3 Consultas Visuais criadas

Foram criadas 4 consultas visuais (views), objetivando a criação de fluxos ETLs, para as seguintes saídas:

- Dimensão Customers (dCustomers);
- Dimensão Employee (dEmployee)
- Dimensão Product (dProduct)
- Fato Orders (fOrders)

DW\_Azure\_Importers

Página Inicial

Criar

Procurar

Hub de dados do...

Aplicativos

Métricas

Workspaces

Página Inicial

Obter dados | No SQL

Explorer

+ Armazéns

sys

Warehouse

Security

Queries

Consultas compartilhadas

dCustomers

dEmployee

dProduto

fOrders

Figura 44 – Consultas Visuais

The screenshot shows the Power BI desktop interface with the 'dCustomers' query selected in the Explorer pane. The visual is a table with the following columns:

CustomerID	CustomerName	PostalCode	AccountOpenedDate	CustomerCategoryName	BuyingGroupName
1	Tarpin Toys (Head Office)	90410	01/01/2013	Novelties Shop	Tarpin Toys
2	Tarpin Toys (Sydney, MT)	90216	01/01/2013	Novelties Shop	Tarpin Toys
3	Tarpin Toys (Peoples Valley, AZ)	90205	01/01/2013	Novelties Shop	Tarpin Toys
4	Tarpin Toys (Medicine Lodge, KS)	90152	01/01/2013	Novelties Shop	Tarpin Toys
5	Tarpin Toys (Gasport, NY)	90261	01/01/2013	Novelties Shop	Tarpin Toys
6	Tarpin Toys (Leslie, ND)	90298	01/01/2013	Novelties Shop	Tarpin Toys
7	Tarpin Toys (Frankewing, TN)	90711	01/01/2013	Novelties Shop	Tarpin Toys
8	Tarpin Toys (Mesa, CO)	90284	01/01/2013	Novelties Shop	Tarpin Toys
9	Tarpin Toys (Inverness, NL)	90129	01/01/2013	Novelties Shop	Tarpin Toys
10	Tarpin Toys (Winnsboro, ND)	90061	01/01/2013	Novelties Shop	Tarpin Toys
11	Tarpin Toys (Default, RA)	90185	01/01/2013	Novelties Shop	Tarpin Toys
12	Tarpin Toys (Bisney, MN)	90504	01/01/2013	Novelties Shop	Tarpin Toys

Figura 45 – Consulta Visual dCustomers

The screenshot shows the Power BI desktop interface with the 'dProduct' query selected in the Explorer pane. The visual is a table with the following columns:

StockItemID	StockItemName	SupplierID	ColorID	Brand	Size	LeadInDays	Suppliers
1	US\$ food fan drive - green	12	null	null	14	[Table]	
2	US\$ food fan drive - blue	12	12	null	14	[Table]	
3	Office cube penknife (black)	12	3	null	14	[Table]	
4	US\$ food fan drive - sushi roll	12	12	null	14	[Table]	
5	US\$ food fan drive - hamburger	12	null	null	14	[Table]	
6	US\$ food fan drive - hot dog	12	null	null	14	[Table]	
7	US\$ food fan drive - pizza slice	12	null	null	14	[Table]	
8	US\$ food fan drive - dum sum 10 drive variety pack	12	null	null	14	[Table]	
9	US\$ food fan drive - banana	12	null	null	14	[Table]	
10	US\$ food fan drive - chocolate bar	12	null	null	14	[Table]	
11	US\$ food fan drive - cookie	12	null	null	14	[Table]	
12	US\$ food fan drive - donut	12	null	null	14	[Table]	

Figura 46 – Consulta Visual dProduct

The screenshot shows the Power BI desktop interface. On the left, the 'Explorer' pane is open, displaying a tree structure of data sources and queries. Under 'Queries', the 'dEmployee' query is selected. In the main workspace, a table visualization titled 'dEmployee' is displayed, showing columns for PersonID, FullName, and PreferredName. The data includes rows for various employees like Kyla Woodcock, Hudson Olslow, Isabella Rupp, etc.

Figura 47 – Consulta Visual dEmployee

The screenshot shows the Power BI desktop interface. On the left, the 'Explorer' pane is open, displaying a tree structure of data sources and queries. Under 'Queries', the 'fOrders' query is selected. In the main workspace, a table visualization titled 'fOrders' is displayed, showing columns for OrderID, CustomerID, SalespersonPersonID, OrderDate, ExpectedDeliveryDate, OrderLineID, StockItemID, Quantity, UnitPrice, and TaxRate. The data includes rows for various orders placed by different customers on different dates.

Figura 48 – Consulta Visual fOrders

#### 6 . 3 . 4 C r i a ç ã o d e V i e w s

Para que os fluxos fiquem disponíveis para gravação em outros locais, como Lakehouses ou até mesmo banco de dados, foram criadas views com as transformações realizadas.

Por enquanto, o Data Warehouse (Fabric) não possui opção de saída (gravação direta)

Para criar as views, é possível obter o script criado, automaticamente pelo fluxo visual, através do botão “Exibir SQL”

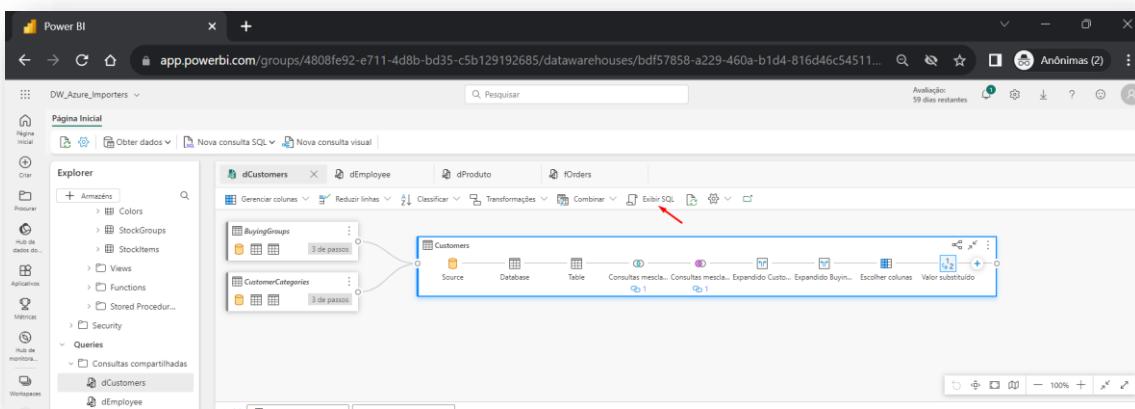


Figura 49 – Exibir script SQL

A screenshot of the 'Exibir SQL' dialog box. It contains a large block of T-SQL code that defines a complex view or query. The code includes multiple SELECT statements, JOIN clauses, and various data transformations. At the bottom right of the dialog, there are two buttons: 'Editar script SQL' and 'Fechar'.

Figura 50 – Exemplo de um script sendo exibido

Assim, ao copiar o Script, pode-se criar uma nova consulta SQL e inserir os termos de criação de view ou escolher a opção “Exibição”, que criará um script de view, com base no resultado do fluxo.

Figura 51 – Nova consulta SQL

```

CREATE VIEW [vw_Orders]
AS
select [$Outer].[OrderID],
[$Outer].[CustomerID],
[$Outer].[SalespersonPersonID],
[$Outer].[OrderDate],
[$Outer].[ExpectedDeliveryDate],
[$Outer].[OrderLineID],
[$Outer].[StockItemID],
[$Outer].[Quantity],
[$Outer].[UnitPrice],
[$Outer].[TaxRate]
from [DW_Azure_Importers].[Sales].[Orders] as [$Outer]
left outer join
(
    select [...].[OrderLineID] as [OrderLineID],
    [...].[OrderID] as [OrderID],
    [...].[StockItemID] as [StockItemID],
    [...].[UnitPrice] as [UnitPrice],
    [...].[TaxRate] as [TaxRate]
    from [DW_Azure_Importers].[Sales].[OrderLines] as [...]
) as [$Inner] on ($Outer).[OrderID] = [$Inner].[OrderID] or ($Outer).[OrderID] is null and [$Inner].[OrderID] is null

```

Figura 52 – Exemplo de script SQL para criação de View

The screenshot shows the Power BI Data Explorer interface. On the left, there's a tree view of databases, tables, and views. A red box highlights the 'Views' node under the 'dbo' schema. The main area displays a table titled 'Visualização de dados' (Data View) with 26 rows of data. The columns are: OrderID, CustomerID, SalespersonPersonID, OrderDate, ExpectedDeliveryDate, OrderLineID, StockItemID, Quantity, UnitPrice, and TaxRate. The table shows various orders placed by different customers on different dates, with their expected delivery dates, line items, stock items, quantities, unit prices, and tax rates.

	OrderID	CustomerID	SalespersonPersonID	OrderDate	ExpectedDeliveryDate	OrderLineID	StockItemID	Quantity	UnitPrice	TaxRate
1	45	892	2	2013-01-01	2013-01-02	1	164	50	112,00	15,000
2	45	892	2	2013-01-01	2013-01-02	2	67	10	230,00	15,000
3	1	892	2	2013-01-01	2013-01-02	2	67	10	230,00	15,000
4	1	892	2	2013-01-01	2013-01-02	3	50	9	13,00	15,000
5	2	893	8	2013-01-01	2013-01-02	3	50	9	13,00	15,000
6	2	893	8	2013-01-01	2013-01-02	4	89	72	18,00	15,000
7	46	893	8	2013-01-01	2013-01-02	4	89	72	18,00	15,000
8	46	893	8	2013-01-01	2013-01-02	5	171	90	32,00	15,000
9	46	893	8	2013-01-01	2013-01-02	5	171	90	32,00	15,000
10	46	893	8	2013-01-01	2013-01-02	6	10	9	32,00	15,000
11	2	893	8	2013-01-01	2013-01-02	6	10	9	32,00	15,000
12	2	893	8	2013-01-01	2013-01-02	7	167	50	99,00	15,000
13	47	105	7	2013-01-01	2013-01-02	7	167	50	99,00	15,000
14	47	105	7	2013-01-01	2013-01-02	8	219	80	50,00	15,000
15	47	105	7	2013-01-01	2013-01-02	8	219	80	50,00	15,000
16	47	105	7	2013-01-01	2013-01-02	9	114	3	30,00	15,000
17	3	105	7	2013-01-01	2013-01-02	9	114	3	30,00	15,000
18	3	105	7	2013-01-01	2013-01-02	10	72	9	230,00	15,000
19	47	105	7	2013-01-01	2013-01-02	10	72	9	230,00	15,000
20	47	105	7	2013-01-01	2013-01-02	11	206	96	2,70	15,000
21	4	57	16	2013-01-01	2013-01-02	12	130	5	32,00	15,000
22	4	57	16	2013-01-01	2013-01-02	12	130	5	32,00	15,000
23	4	57	16	2013-01-01	2013-01-02	13	50	2	13,00	15,000
24	4	57	16	2013-01-01	2013-01-02	13	50	2	13,00	15,000
25	4	57	16	2013-01-01	2013-01-02	13	50	2	13,00	15,000
26	4	57	16	2013-01-01	2013-01-02	13	50	2	13,00	15,000

Bem-sucedido (1 sec 198 ms)

Figura 53 – Lista de Views criadas

Se o objetivo final, nessas etapas de teste, fosse a criação de relatórios de análise em Power BI, o processo de ETL poderia encerrar aqui.

A criação de relatórios, pode ser realizada, tanto no ambiente de Warehouse quanto criando um novo relatório e escolhendo a fonte de dados Warehouse.

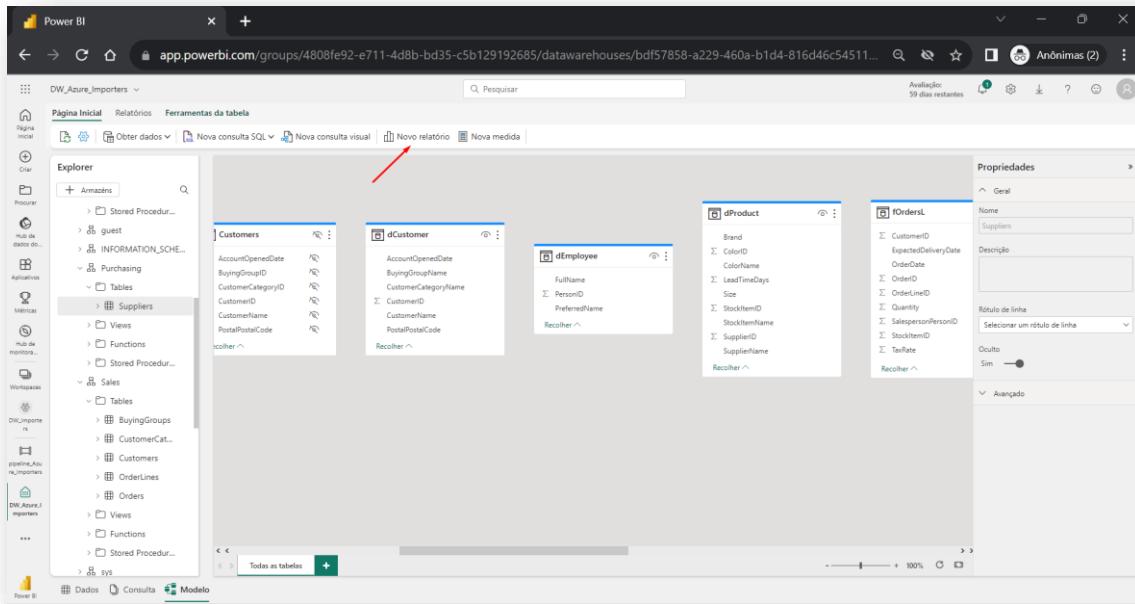


Figura 54 – Opção “Modelo” – com visuais (ativos) disponíveis para conexão pelo PBI

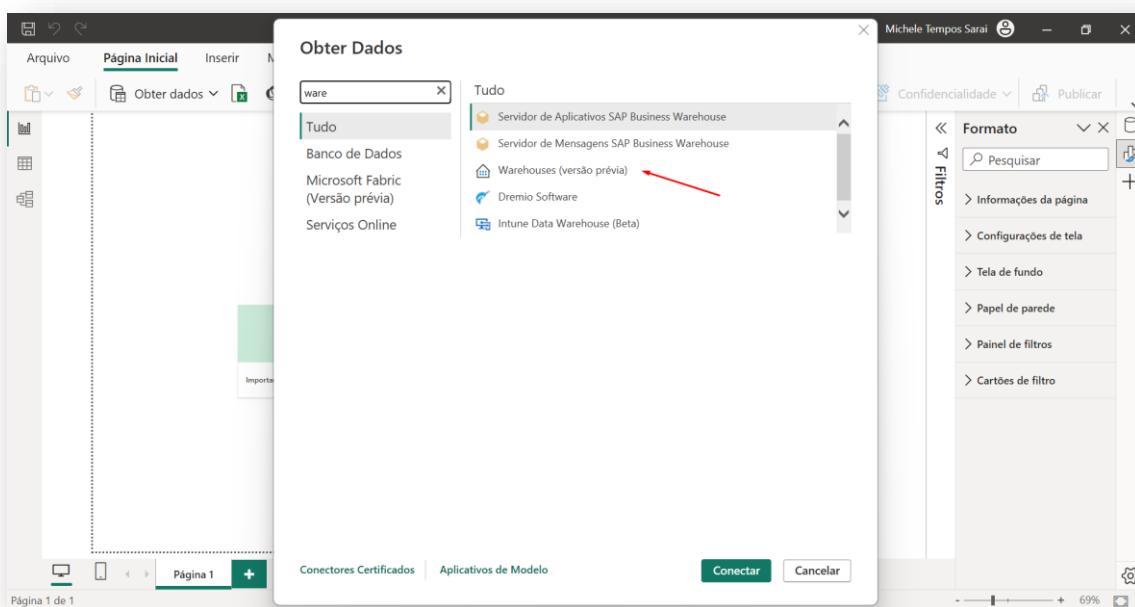


Figura 55 – Power BI – Obtenção de Dados – conectar Warehouse

## 6 . 3 . 5 Gravação dos dados em um novo banco de dados

Como o objetivo é formar um banco de dados com o resultado do ETL e, como o Warehouse não possui essa opção, ainda, então, houve a necessidade de criar um Fluxo de dados (Data Factory) para coletar as views, presentes no Warehouse e destinar para um banco.

Uma observação importante, é que todo processo de ETL poderia ter sido feito pelo Fluxo de dados, entretanto, a ideia inicial era testar o Warehouse, do Fabric.

Nos prints seguintes, realizei, inclusive alguns pontos de ETL, apenas para exemplificar o uso no Fluxo de Dados e, até para deixar claro, que um processo pode complementar o outro.

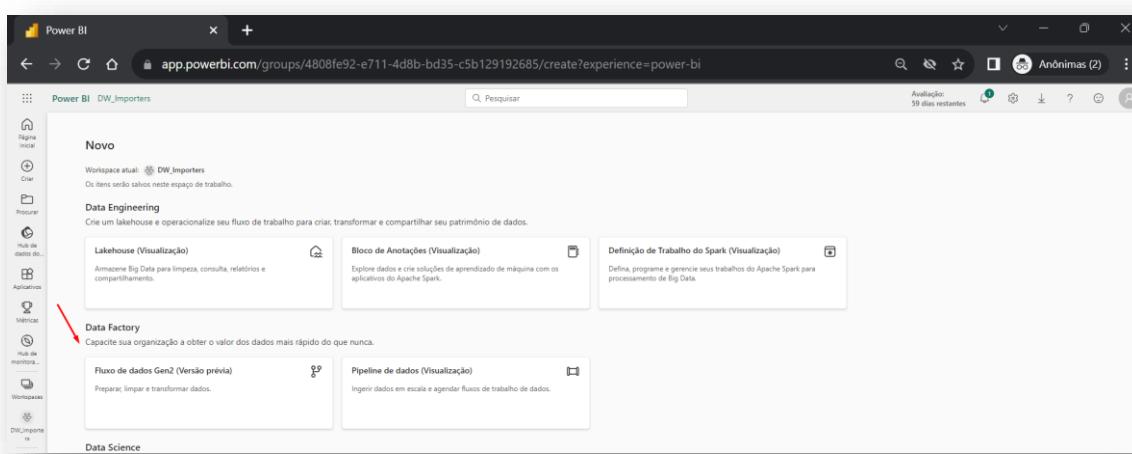


Figura 56 – Fluxo de Dados (Data Factory)

A screenshot of a web browser showing the Power BI workspace interface. The URL is app.powerbi.com/groups/4808fe92-e711-4d8b-bd35-c5b129192685/list?experience=power-bi&\_iframeld=. The workspace name is DW\_Importers. The sidebar has icons for Página Inicial, Criar, Procurar, and Hub de. The main area shows a table of data flows. The table has columns: Nome, Tipo, Proprietário, Atualizado, Próxima atualização, and Endoso. One row is visible: 'Dataflow DW from RDS' (Nome), 'Fluxo de dados Gen2' (Tipo), 'Michele Tempos Sarai' (Proprietário), '07/09/23, 09:48:38' (Atualizado), 'N/D' (Próxima atualização), and an empty field (Endoso). A red arrow points from the 'Nome' column towards the 'Dataflow DW from RDS' entry.

Figura 57 – Fluxo de Dados criado (Dataflow)

Lembrando que o Data Factory possui diversos conectores de dados, inclusive com o Warehouse

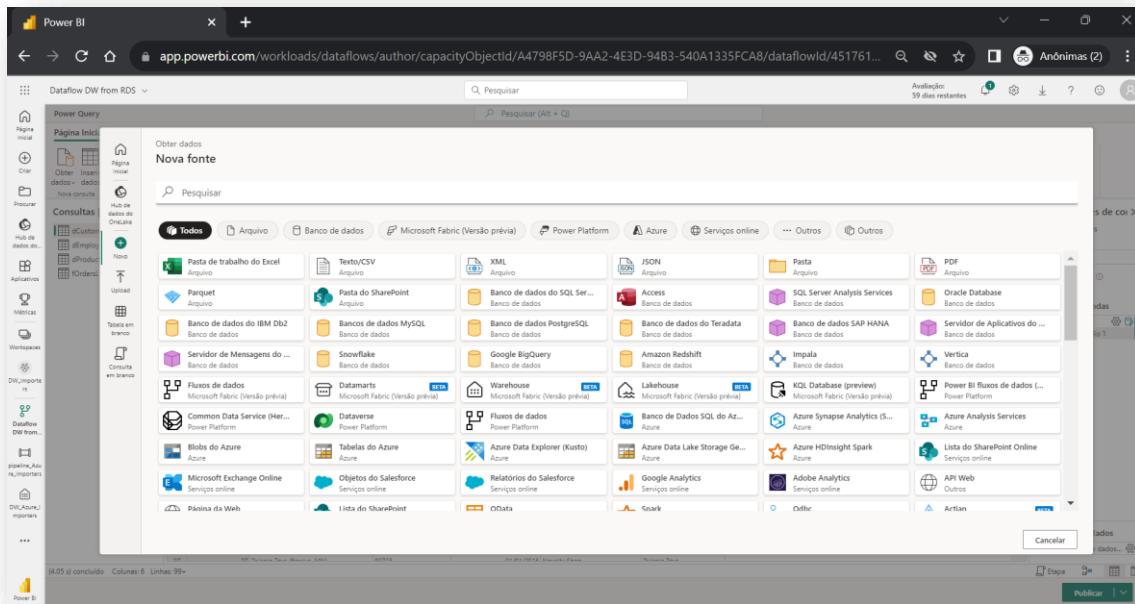


Figura 58 – escolha da fonte dados no Dataflow

Abaixo, o print demonstra que é possível criar as etapas de processamento do ETL dentro do Dataflow. No exemplo, foram removidos os valores duplicados, existentes na consulta dProduct, assim como também foi acrescentada uma etapa de substituição de valores nulos por um conteúdo específico.

StockItemID	SupplierName	SupplierID	Brand	Size	LastUnitBought	SupplierName	ColorName
1	USB inside launcher (Green)	12	null	Undefined	null	The Phone Company	no color
2	USB rocket launcher (Grey)	12	12	Undefined	null	The Phone Company	Steel Gray
3	Office cube periscope (Black)	12	3	Undefined	null	The Phone Company	Black
4	USB food flash drive - small	12	12	Undefined	null	The Phone Company	no color
5	USB food Flash drive - hamberger	12	null	Undefined	null	The Phone Company	no color
6	USB food Flash drive - hot dog	12	null	Undefined	null	The Phone Company	no color
7	USB food Flash drive - pizza slice	12	null	Undefined	null	The Phone Company	no color
8	USB food Flash drive - dim sum 10 drive variety pack	12	null	Undefined	null	The Phone Company	no color
9	USB food Flash drive - banana	12	null	Undefined	null	The Phone Company	no color
10	USB food Flash drive - chocolate bar	12	null	Undefined	null	The Phone Company	no color
11	USB food Flash drive - cookie	12	null	Undefined	null	The Phone Company	no color
12	USB food Flash drive - donut	12	null	Undefined	null	The Phone Company	no color
13	USB food Flash drive - coffee cocktail	12	null	Undefined	null	The Phone Company	no color
14	USB food Flash drive - fortune cookie	12	null	Undefined	null	The Phone Company	no color
15	USB food Flash drive - dessert 10 drive variety pack	12	null	Undefined	null	The Phone Company	no color
16	DGA joke mug - mind it! join yeul! (White)	3	33	Undefined	null	Graphic Design Institute	White
17	DGA joke mug - mind it! join yeul! (White)	3	3	Undefined	null	Graphic Design Institute	Black
18	DGA joke mug - daaaaaa-ta (White)	3	33	Undefined	null	Graphic Design Institute	White
19	DGA joke mug - daaaaaa-ta (Black)	3	3	Undefined	null	Graphic Design Institute	Black
20	DGA joke mug - you might be a DBA if (White)	3	33	Undefined	null	Graphic Design Institute	White
21	DGA joke mug - you might be a DBA if (Black)	3	3	Undefined	null	Graphic Design Institute	Black
22	DGA joke mug - you might be a DBA if (White)	3	33	Undefined	null	Graphic Design Institute	Black
23	DGA joke mug - i deserve it! (Black)	3	3	Undefined	null	Graphic Design Institute	Black
24	DGA joke mug - i will get you in order (White)	3	33	Undefined	null	Graphic Design Institute	White
25	DGA joke mug - i will get you in order (Black)	3	3	Undefined	null	Graphic Design Institute	Black
26	DGA joke mug - SELECT caffeine FROM mug (White)	3	33	Undefined	null	Graphic Design Institute	White
...	...	...	...	...	...	...	...
99	...	...	...	...	...	...	...

Figura 59 – Exemplo de processo de ETL

Para a gravação em um banco de dados, deve-se adicionar um Destino de Dados e em seguida, Publicar. Para o destino final, foi criado um banco de dados, denominado dwImportes

Figura 60 – Escolha do Destino de gravação

Figura 61 – Publicação do conteúdo

Para demonstrar a criação das tabelas, veja, a seguir, os prints de conexão ao banco.

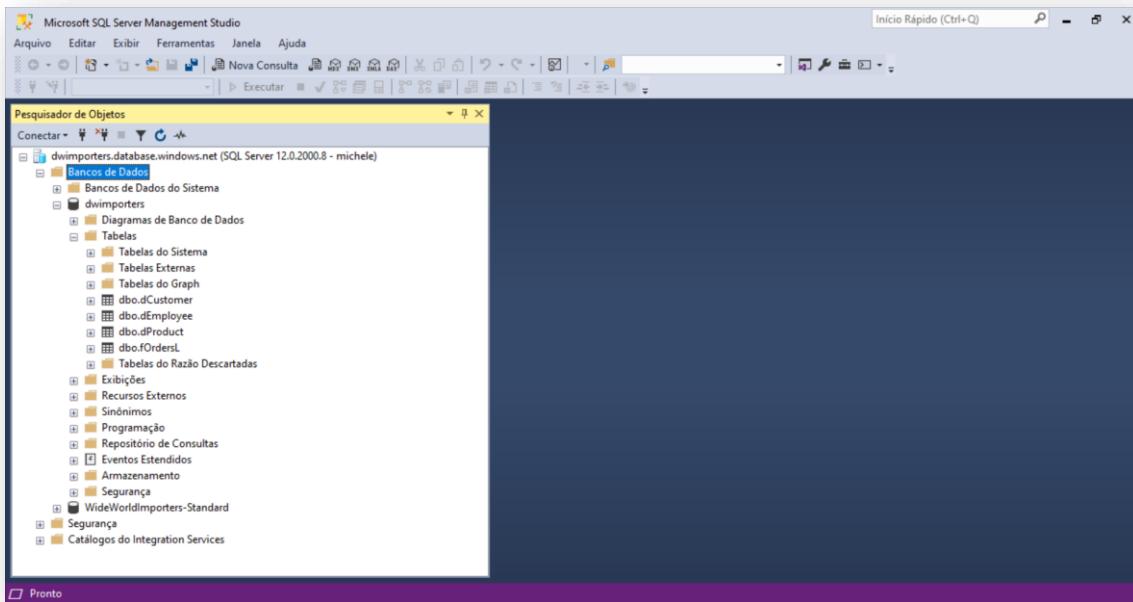


Figura 62 – Conexão via SSM

The screenshot shows the Microsoft Azure portal interface. The address bar indicates a connection to "dwimporters (dwimporters/dwimporters) | Editor de consultas (visualização)". The left sidebar contains navigation links for "Visão geral", "Log de atividade", "Marcações", "Diagnosticar e resolver problemas", "Editor de consultas (visualização)", "Configurações" (with options for "Computação + armazenamento", "Cadeias de conexão", "Propriedades", and "Bloqueios"), and "Gerenciamento de dados" (with "Réplicas"). The main area shows the "dwimporters (michele)" database structure, with "Tabelas" expanded to show "dbo.dCustomer", "dbo.dEmployee", "dbo.dProduct", and "dbo.OrdersL". Below this, "Procedimentos Armazenados" and "Exibições" are listed. A "Consulta 2" tab is active, containing a query editor with the following code:

```
1 SELECT TOP (1000) * FROM [dbo].[dCustomer]
```

The "Resultados" tab is selected, showing a table with two rows of data:

CustomerID	CustomerName	PostalCode	AccountOpenedDate	CustomerType
1001	Dinh Mai	90423	2013-01-27T00:00:00...	Computer
1002	Stefan Selezeanu	90053	2013-02-22T00:00:00...	Computer

A message at the bottom of the results table says "Éxito na consulta | 0s".

Figura 63 – Conexão via Editor de consultas na Azure

## 7. ANÁLISE

A qualidade de dados é fundamental em qualquer organização, pois afeta diretamente a tomada de decisões, a eficiência operacional e a confiabilidade das informações. Ter dados de alta qualidade é crucial para a obtenção de insights precisos e confiáveis, o que, por sua vez, impulsiona o sucesso dos negócios.

Existem diversas formas de estabelecer a qualidade de dados:

1. Coleta de dados precisa: Garanta que os dados sejam coletados de forma precisa desde o início. Isso envolve a implementação de formulários e sistemas de registro eficientes.
2. Padronização e normalização: Utilize padrões e convenções para uniformizar os dados, tornando-os consistentes e comparáveis.
3. Limpeza de dados: Regularmente, faça uma limpeza dos dados para eliminar duplicações, erros de digitação e informações incompletas ou inconsistentes.
4. Validação e verificação: Implemente processos de validação e verificação para garantir que os dados estejam corretos e atendam aos requisitos estabelecidos.
5. Governança de dados: Estabeleça políticas e responsabilidades claras para o gerenciamento de dados, incluindo quem é responsável pela qualidade dos dados.
6. Monitoramento contínuo: Mantenha um monitoramento constante da qualidade dos dados para identificar e corrigir problemas rapidamente.
7. Treinamento e conscientização: Eduque os funcionários sobre a importância da qualidade de dados e forneça treinamento adequado para garantir que eles saibam como coletar e gerenciar os dados de forma eficaz.
8. Uso de tecnologia: Utilize ferramentas de qualidade de dados, como software de limpeza e integração, para automatizar processos e melhorar a consistência dos dados.

Durante o processo de transformação de dados, foram realizadas verificações sobre a qualidade dos dados. Primeiramente, serão demonstrados alguns problemas encontrados, em seguida, a resolução.

## 7.1 PROBLEMAS

### A) Produtos duplicados

Havia registros duplicados para produtos. A comprovação pode ser realizada de duas formas:

- Script de contagem;
- Qualidade, distribuição e perfil da coluna, no dataflow (data factory)

StockItemID	StockItemName	contagem
9	USB food flash drive - banana	4
12	USB food flash drive - donut	4
24	DBA joke mug - I will get you in order (White)	8
28	DBA joke mug - two types of DBAs (Black)	8
34	Developer joke mug - fun was unexpected at this time (White)	8
77	'The Gu' red shirt XML tag t-shirt (White) XS	8
80	'The Gu' red shirt XML tag t-shirt (White) M	8
86	'The Gu' red shirt XML tag t-shirt (White) SXL	8
109	Superhero action jacket (Blue) XS	8
116	Superhero action jacket (Blue) XL	8
130	Furry gorilla with big eyes slippers (Black) S	8
132	Furry gorilla with big eyes slippers (Black) L	8
133	Furry gorilla with big eyes slippers (Black) XL	8
136	Furry animal socks (Pink) S	4
142	Halloween zombie mask (Light Brown) S	8

Figura 64 - script de contagem de registros

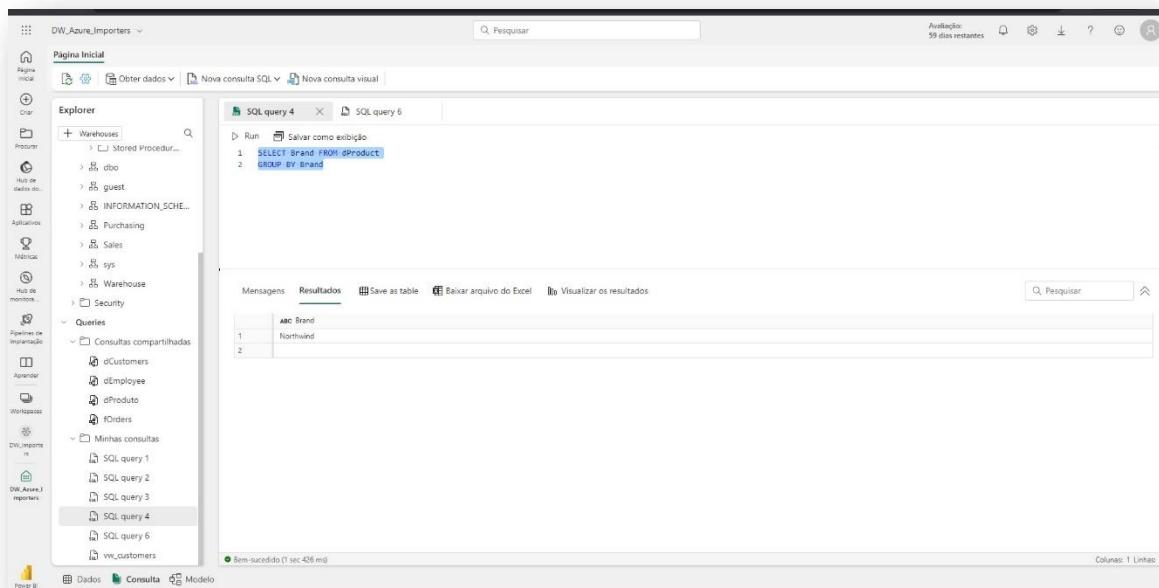
Figura 65 - Exibições do perfil da coluna

## B) Produtos sem marca (brand)

Vários registros estavam sem definição da marca. Entretanto, como um dos pontos de análise é a identificação de vendas por marca, preencheu-se os registros vazios com um valor default. Entretanto, em uma situação dessa, o ideal é conversar com o responsável pela fonte de dados original.

Para a identificação, utilizou-se:

- Script sql;
- Qualidade, distribuição e perfil da coluna, no dataflow (data factory)



The screenshot shows the Microsoft Data Factory interface. On the left, there's a navigation pane with various options like 'Página inicial', 'Explorador', 'Consultas', 'Relatórios', and 'DW\_Imports'. In the center, there's a 'SQL query 4' tab with the following SQL code:

```
SELECT Brand FROM dproduct
GROUP BY Brand
```

The results pane shows the output of the query:

Brand
Northwind

At the bottom, it says 'Bem-sucedido (1 sec 406 ms)' and 'Column: 1 Linhas: 2'.

Figura 66 - Script SQL para identificação das marcas do produtos

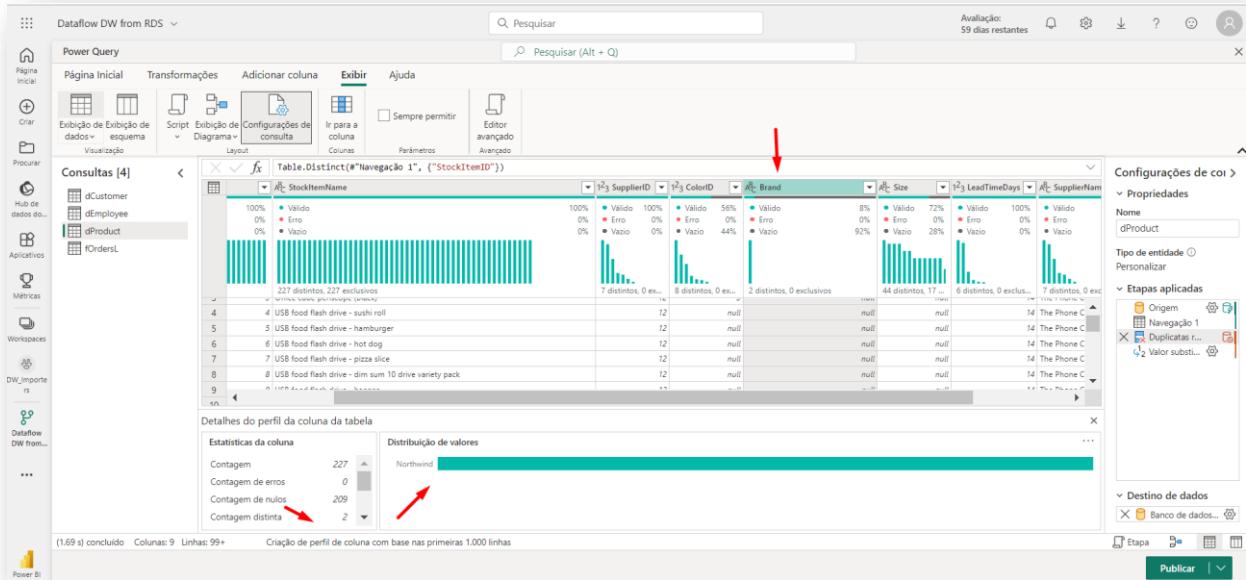


Figura 67 - Perfil da coluna

### C) Clientes - BuyingGroupName

Vários registros estavam sem definição da grupo de compra. Entretanto, como um dos pontos de análise é a identificação de vendas por marca, preencheu-se os registros vazios com um valor default. Entretanto, em uma situação dessa, o ideal é conversar com o responsáveis pela fonte de dados original.

Para a identificação, utilizou-se:

- Qualidade, distribuição e perfil da coluna, no processo de ETL do Data Warehouse

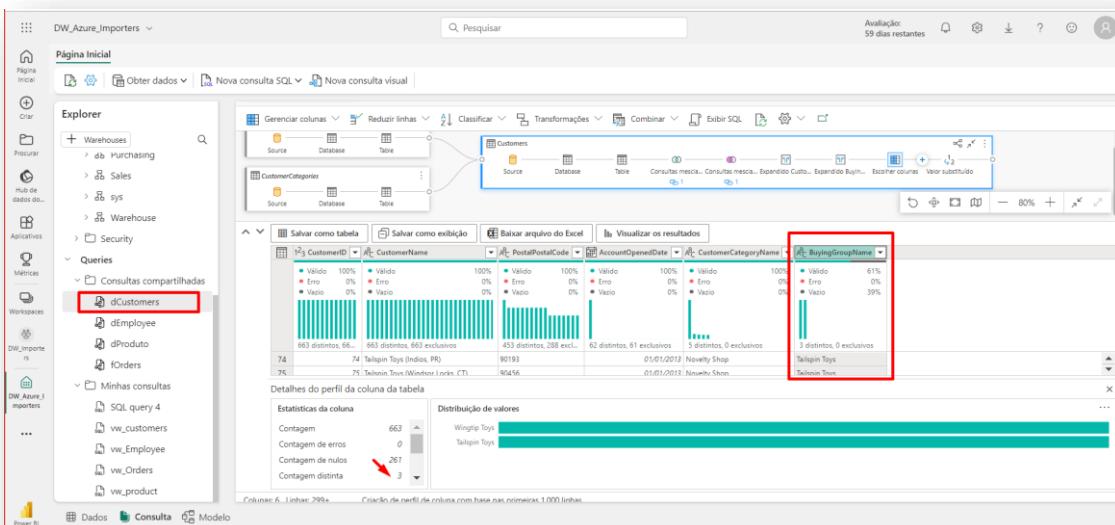


Figura 68 - BuyingGroupName nulos

## 7.2 RESOLUÇÃO

### A) Produtos duplicados

No dataflow, foi acrescentada uma etapa para remover as duplicatas

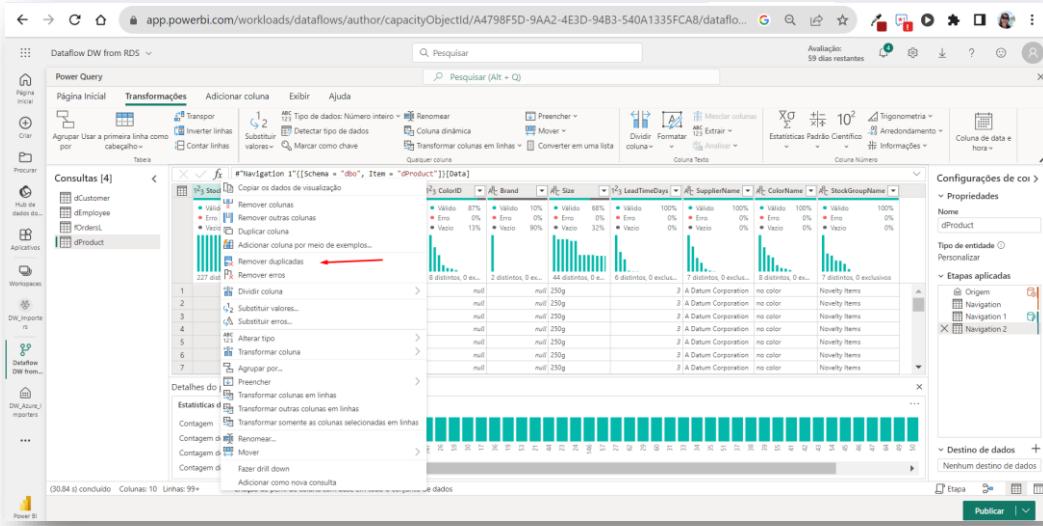


Figura 69 - Acrescentando componente para remoção de duplicatas

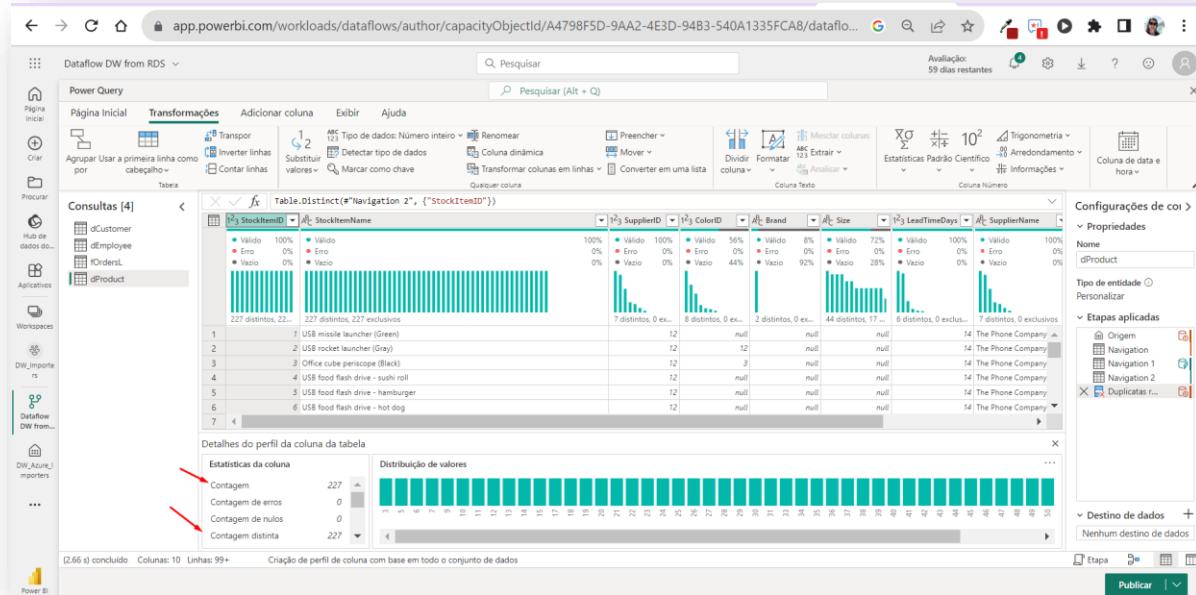


Figura 70 - Indicador de contagem de registros e exclusivos

## B) Produtos sem marcas (brand)

No dataflow, foi acrescentada uma etapa para substituir valores nulos

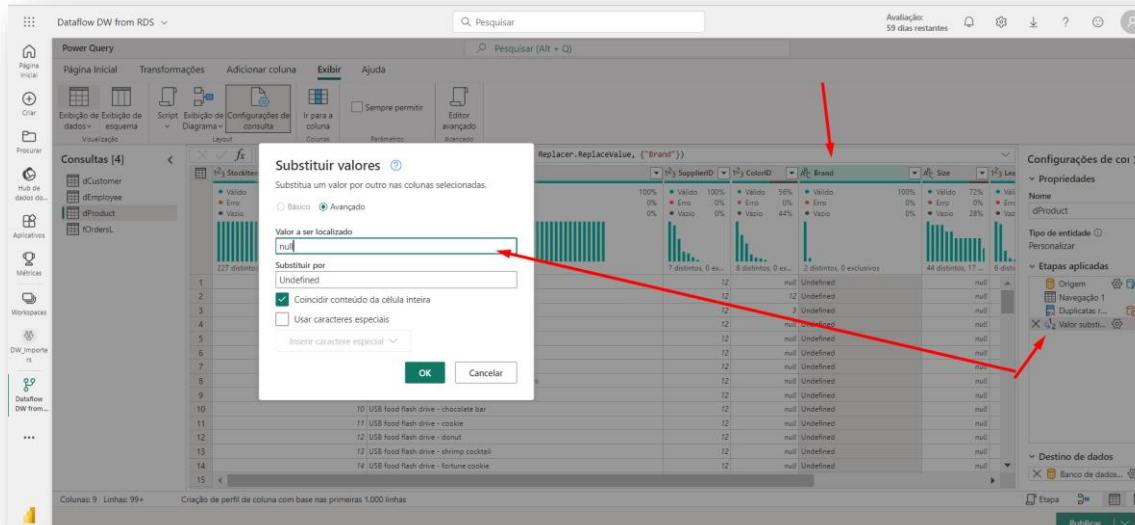


Figura 71 - Inclusão de componente de substituição de valores

## C) BuyingGroupName – Nulos

No ETL, do Data Warehouse, acrescentou-se um componente de substituição de valores

Figura 72 - Componente de substituição de valores

# 8. ANÁLISE DOS OBJETIVOS ESTABELECIDOS

No início deste projeto, foram especificados algumas questões analíticas.

Para apresentar os resultados, escolheu-se a ferramenta de Data Viz, Power BI.

## 1. Dos quatro anos disponíveis para análise, qual foi o melhor ano, em termos de faturamento?

O melhor ano, em termos de valor faturado, foi o de 2016, com \$ 55.817.887,45. O visual gráfico destaque o ano com maior venda, observe



Figura 73 - Ano com maior valor faturado

## 2. Do catálogo de produtos, há algum produto que não possuí vendas?

Nos dados atuais, não há a data de cadastro do produto. Portanto, a análise de contagem do catálogo está considerando que são 227 produtos disponíveis para venda e que, estavam disponíveis em todos os anos.

Este é um ponto de melhoria, já que a análise de proporção de venda de catálogo, por exemplo, tem que levar em consideração apenas os existentes na época.

O catálogo de produtos, possuí 227 cadastros, sendo que no último ano, todos os produtos foram negociados.

Anos que alguns produtos não foram negociados:

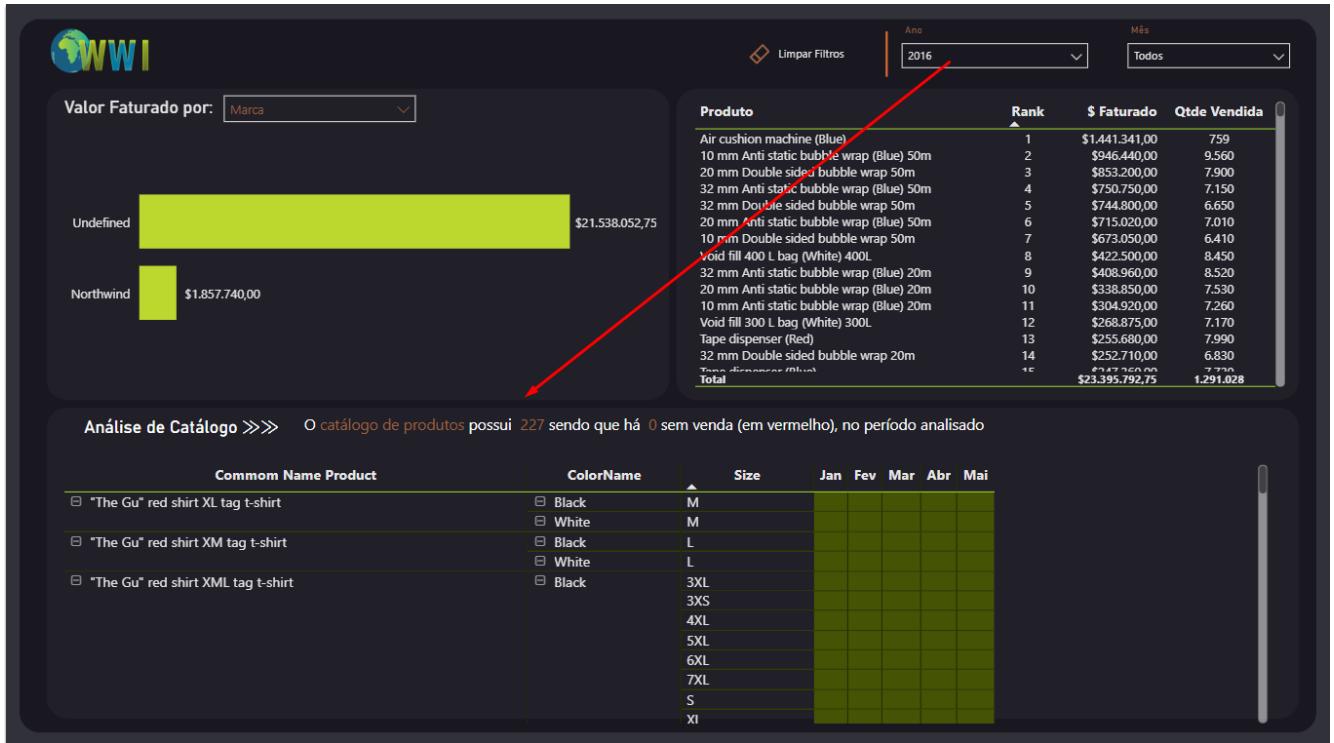


Figura 74 - Catálogo de dados 100% vendido em 2016

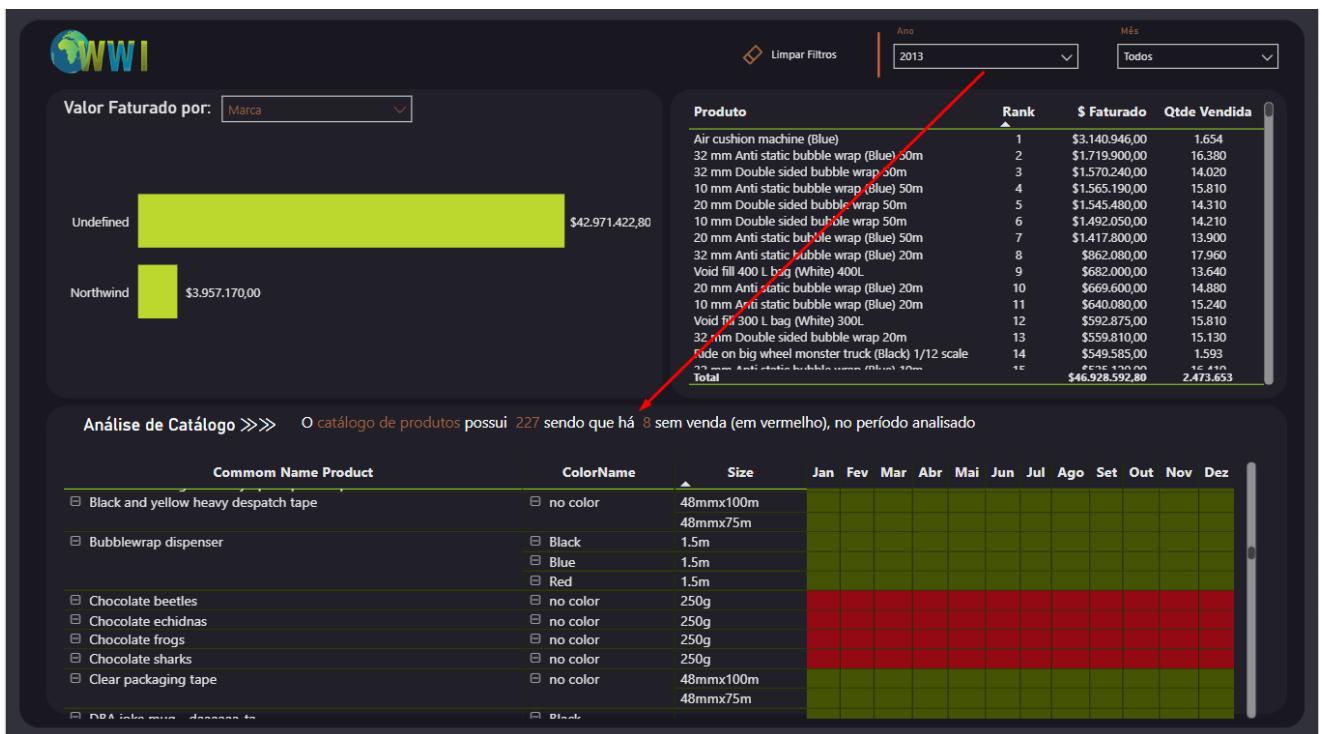


Figura 75 - Catálogo de dados com 8 produtos não vendidos em 2013

### 3. No último ano, como estão as vendas? Há variação positiva?

Temos que observar que o último ano possui vendas até o período do mês de Maio. Dessa forma, em uma primeira análise, parece que o Ano de 2016 não está bem, em termos de valor faturado. Entretanto, temos que comparar períodos iguais.

Quando levamos esse quesito em consideração, notamos que as vendas estão positivas, tanto para valores quanto para quantidades, tanto no visual de evolução mensal, quanto no visual de valor acumulado no Ano.

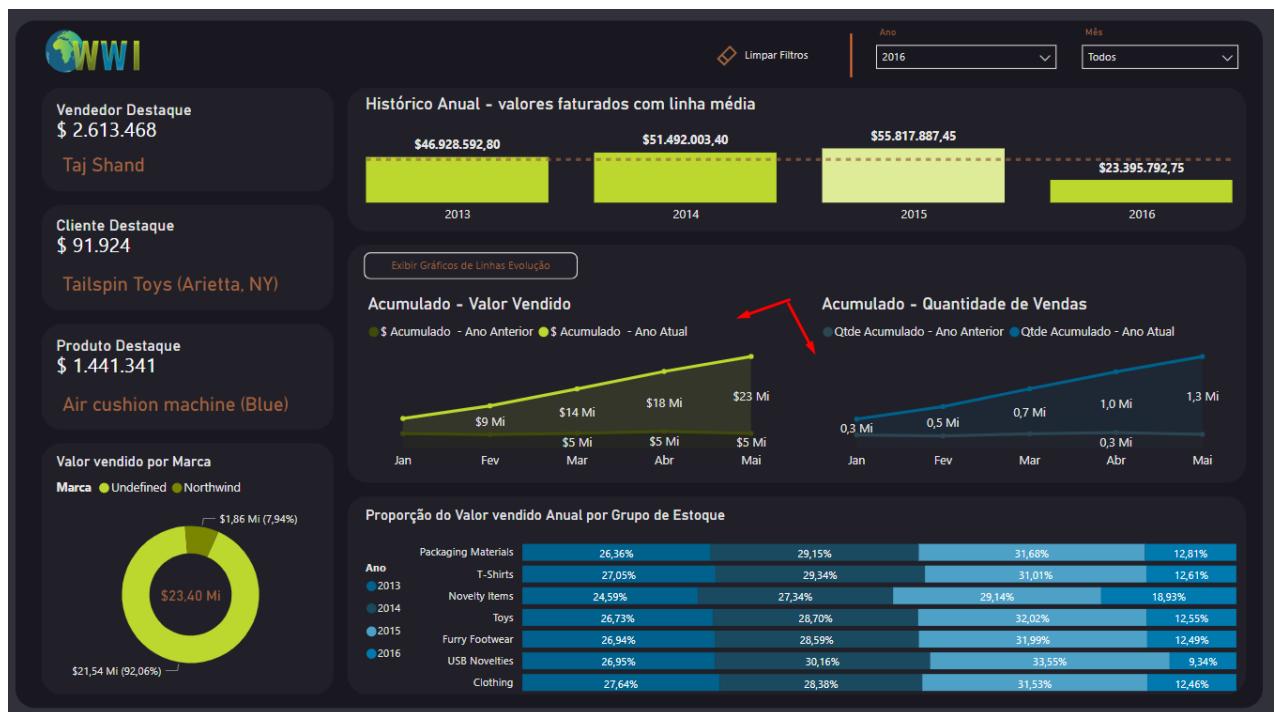


Figura 76 - Desempenho em vendas - Acumulado

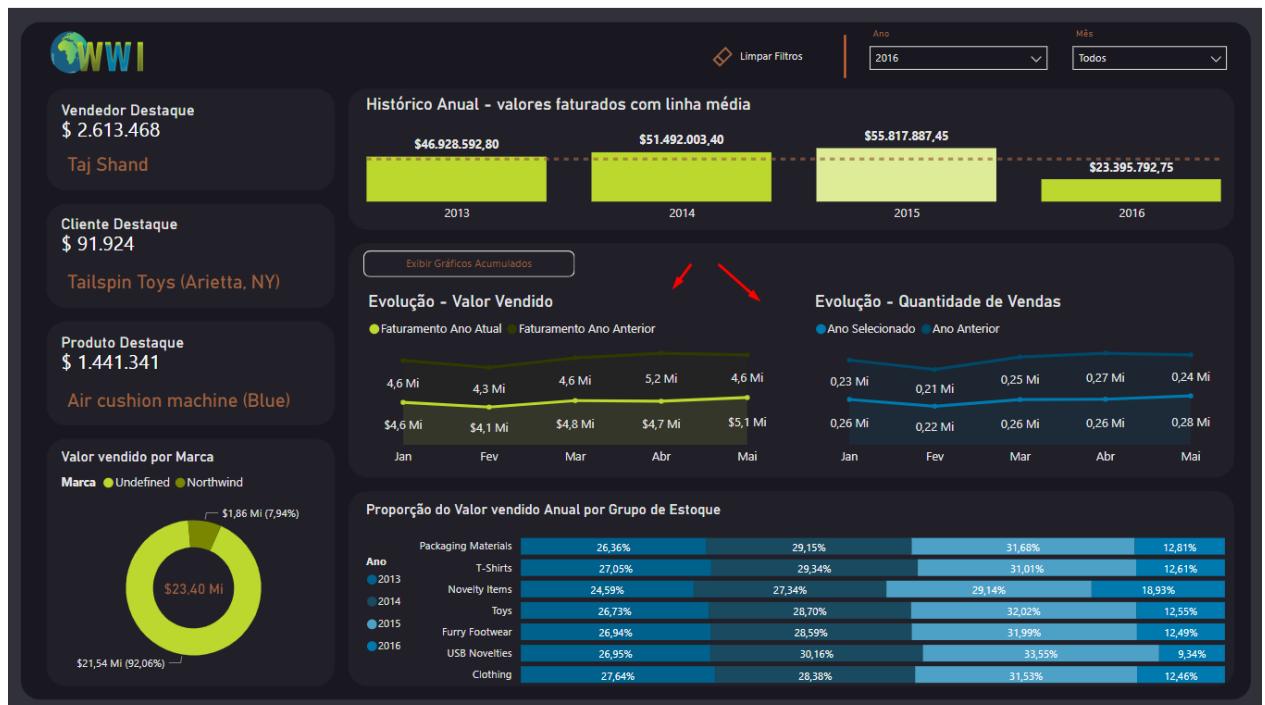


Figura 77 - Desempenho em vendas - Evolução

#### 4. Qual é o cliente que mais faturou?

Podemos fazer a análise de várias maneiras, por categoria de produtos, por Ano, por meses. Como não foi especificado, considerou-se a somatória dos 4 anos, sendo que o melhor cliente foi **Tailspin Toys (Inguadona, MN), com \$ 384.393,35**



Figura 78 - Ranking de Cliente

## 5. No último ano, quantos produtos representam 80% do meu faturamento?

Um total de 70 produtos compreende até 80% do faturamento total, correspondendo a 30% do catálogo completo. Isso sugere uma notável diversidade de produtos que despertam o interesse da nossa carteira ativa de clientes.



Figura 79 - Gráfico de Pareto - Distribuição de produtos por valor faturado

## **6. Em quais produtos / grupos possuo maior margem bruta?**

Esta pergunta não pode ser respondida, neste momento, já que os dados disponibilizados não possuem informações de custo. Um ponto de forte atenção, para que Margens e Lucros possam ser estabelecidos.

## **7. Qual é a projeção para os próximos meses?**

Neste momento, não foi possível estabelecer a projeção, já que demandará um alinhamento dos requisitos/objetivos, como :

- a) Estabelecer o período de base em dados históricos
- b) Analisar tendências de mercado e outros fatores relevantes.

Link para acesso aos relatórios desenvovidos

<https://app.powerbi.com/view?r=eyJrIjoiNzU5NjQ2ZmYtMzdiNi00ZmJkLWFmMGUtMGEyNmUxYTY0OTIiIwidCI6IjE3ZjMzMzMDk3LWJiNjQtNDY2OC05MDYwLWFiMWE2MzgxMmE0MyJ9&pageName=ReportSection>

Obs.: Para a transição de páginas, utilize as setas centrais na parte inferior do relatório. No layout aplicativo, disponível apenas dentro do WorkSpace, a publicação cria, automaticamente, um menu lateral. Esse é o motivo de não terem sido criados botões para transição, focando no aproveitamento dos espaços, para a construção de visuais.