

# MVP3 - Engenharia de Dados

Aluno: José Roberto Assis Silva

RA: 4052023000345

---

## 1. Objetivo

Foi criada uma campanha para premiar os vendedores de equipamentos na empresa em que trabalho. Vou tratar esta campanha neste MVP coletando dados do nosso ERP que utiliza um banco de dados relacional (SGDB ORACLE). O objetivo desta campanha é rankear os vendedores por total de equipamentos vendidos mostrando também a quantidade vendida por grupo de equipamentos. A partir destes dados será feita uma premiação para o primeiro, segundo e terceiro colocados. Esta campanha utilizará apenas dados do ano de 2023, mas vou gerar os CSVs com todas as informações de vendas que estão em nosso database para criar meu dataset. As visualizações destas informações deveram estar em um dashboard mostrando um grid rankeando os vendedores pela quantidade total de equipamentos vendidos e deve conter a quantidade vendida de cada grupo de equipamentos com os totais gerais do grupo e do vendedor. Também terá um gráfico de barras rankeando os vendedores com suas respectivas quantidades totais vendidas. Neste dashboard devemos ter filtro por loja e data. A partir do database (Oracle) vou criar algumas queries para gerar alguns arquivos CSVs e vou utilizá-los para carga na nuvem Azure para fazer ETL e posteriormente criar o dashboard. Vou mudar algumas descrições nos CSVs para manter a confidencialidade dos dados. Os CSVs foram gerados a partir de uma conexão feita com o SGDB Oracle utilizando o Dbeaver e queries SQL.

Objetivos:

1. Criar um grid com ranking dos vendedores por quantidade total vendida.
  2. Nas colunas deste grid deveram ter a quantidade por grupo de equipamento vendido, com o total geral de todos os grupos no final das colunas.
  3. O grid terá o total por vendedor nas linhas.
  4. Fazer um gráfico de barras com ranking dos vendedores por total de equipamentos vendidos.
- 

## 2. Busca pelos Dados

Vou utilizar uma conexão com o SGDB Oracle utilizando o Dbeaver para gerar os CSVs que farão parte do meu dataset. Como conhecedor do database que utilizamos, criei algumas queries, com visões temporárias, joins e filtros para gerar um CSV com itens vendidos apenas de EQUIPAMENTOS (VENDAS), porque também trabalhamos com outras linhas de produtos. Vou gerar também CSVs de 2 tabelas dimensões, as quais utilizarei quando criar um dataflow no Azure Data Factory, fazendo alguns joins para gerar somente uma tabela flat VENDAS não **normalizada**. Farei alterações em algumas descrições preservando a confidencialidade dos dados, também inibirei alguns dados nas imagens. Serão gerados os seguintes CSVs:

1. VENDAS.csv (Contém os atributos das vendas)
2. LOJAS.csv (Contém os atributos das lojas)
3. GRUPOSEQP (Contém os atributos dos grupos dos equipamentos)

## 2.1. Executando query e gerando o CSV VENDAS

DBeaver 23.2.0 - <NBS\_TESTE> Script-2

```
File Edit Navigate Search SQL Editor Database Window Help
Database Navigator X
Enter a part of object name here
> NBS_TESTE - NBS
> NIPVIEW - 192.168.20.235:3306
> TRON-PRODUÇÃO - 192.168.20.32:3050
> TRONCJDB - localhost:3050

WITH OPERACAO AS (
    SELECT O.COD_EMPRESA
        , O.COD_OPERACAO
        , O.OPERACAO
        , O.COD_GRUPO_PC
        , O.COD_NATUREZA_RECEITA_DESPESA
        , NRD.DESCRICAO_DESCR_NRD
        , O.GRUPO
        , O.COD_NATUREZA
        , NAT.DESCRICAO AS DESCR_NATUREZA
        , NAT.DESCRICAO_REFERENCIA AS DESCR_NAT_REF
        , O.COD_CONTA_CORRENTE
        , CC.DESCRICAO AS DESCR_CONTA_CORRENTE
        , O.COD_TIPO_PAGAMENTO
        , TP.DESCRICAO AS DESCR_TIPO_PAGAMENTO
    FROM NBS.OPERACOES O
    LEFT JOIN NBS.NATUREZA_RECEITA_DESPESA NRD ON O.COD_GRUPO_PC = NRD.COD_GRUPO_PC
                                                AND O.COD_NATUREZA_RECEITA_DESPESA = NRD.COD_NATUREZA_RECEITA_DESPESA
    LEFT JOIN NBS.NATUREZA NAT ON O.GRUPO = NAT.COD_NATUREZA
                                                AND O.COD_NATUREZA = NAT.COD_NATUREZA
    LEFT JOIN NBS.CONTA_CORRENTE CC ON O.COD_EMPRESA = CC.COD_EMPRESA
                                                AND O.COD_CONTA_CORRENTE = CC.COD_CONTA_CORRENTE
    LEFT JOIN NBS.TIPO_PAGAMENTO TP ON O.COD_EMPRESA = TP.COD_EMPRESA
                                                AND O.COD_TIPO_PAGAMENTO = TP.COD_TIPO_PAGAMENTO
    WHERE O.cod_operacao in (-1, --Peças Balcão
                            --2, --Serviços
                            --3, --Serviços Garantias
                            4, --Máquinas
                            9, --Máquinas
                            --7, --Peças Devolução
                            --13,
                            --19, --Peças Atacado
                            22, --Diversos
                            --26
                            60, --Devolução de Máquinas fins diversos
                            28 --Máquinas Devolução
                            )
)
SELECT TRUNC(CAB.EMISSAO) AS DATA_FATURAMENTO
        , CAB.CONTROLE
        , CAB.SERIE
        , ED.DESCRICAO AS DEPARTAMENTO
        , CAB.VENDEDOR
        , CAB.COD_CLIENTE_NOTA
        , CAB.COD_CLIENTE
        , CAB.COD_OPERACAO
        , CAB.TOTAL_NOTA AS VALOR_VENDIDO
        , CAB.VALOR_ICMS*(-1) AS VALOR_ICMS
        , CAB.ICMS_SUBSTITUIDO
,
```

Pesquisar 30°C Ensolado 11:15 POR PTB2 21/09/2023

DBeaver 23.2.0 - <NBS\_TESTE> Script-2

```
File Edit Navigate Search SQL Editor Database Window Help
Database Navigator X
Enter a part of object name here
> NBS_TESTE - NBS
> NIPVIEW - 192.168.20.235:3306
> TRON-PRODUÇÃO - 192.168.20.32:3050
> TRONCJDB - localhost:3050

)
SELECT TRUNC(CAB.EMISSAO) AS DATA_FATURAMENTO
        , CAB.CONTROLE
        , CAB.SERIE
        , ED.DESCRICAO AS DEPARTAMENTO
        , CAB.VENDEDOR
        , CAB.COD_CLIENTE_NOTA
        , CAB.COD_CLIENTE
        , CAB.COD_OPERACAO
        , CAB.TOTAL_NOTA AS VALOR_VENDIDO
        , CAB.VALOR_ICMS*(-1) AS VALOR_ICMS
        , CAB.ICMS_SUBSTITUIDO
        , CAB.ANO_MODELO
        , CAB.ANO_FABRICACAO
        , CAB.VALOR_PIS*(-1) AS VALOR_PIS
        , CAB.VALOR_COFINS*(-1) AS VALOR_COFINS
        , CAB.STATUS
        , VS.DESCRICAO AS DESCRSTATUS
        , CAB.END_CID_DESCRICAO
        , CAB.END_CID_UF
        , CAB.END_CID_DESCRICAO || '' || CAB.END_CID_UF AS C
--Dados da Fabrica

Results 1 <
WITH OPERACAO AS (SELECT O.COD_EMPRESA , O.COD_OPERACAO , O.OPERACAO , O.COD_GRUPO_PC , O.COD_NATUREZA_RECEITA_DESPESA , NRD.DESCRICAO_DESCR_NRD , O.GRUPO , O.COD_NATUREZA , NAT.DESCRICAO AS DESCR_NATUREZA , NAT.DESCRICAO_REFERENCIA AS DESCR_NAT_REF , O.COD_CONTA_CORRENTE , CC.DESCRICAO AS DESCR_CONTA_CORRENTE , O.COD_TIPO_PAGAMENTO , TP.DESCRICAO AS DESCR_TIPO_PAGAMENTO
    FROM NBS.OPERACOES O
    LEFT JOIN NBS.NATUREZA_RECEITA_DESPESA NRD ON O.COD_GRUPO_PC = NRD.COD_GRUPO_PC
                                                AND O.COD_NATUREZA_RECEITA_DESPESA = NRD.COD_NATUREZA_RECEITA_DESPESA
    LEFT JOIN NBS.NATUREZA NAT ON O.GRUPO = NAT.COD_NATUREZA
                                                AND O.COD_NATUREZA = NAT.COD_NATUREZA
    LEFT JOIN NBS.CONTA_CORRENTE CC ON O.COD_EMPRESA = CC.COD_EMPRESA
                                                AND O.COD_CONTA_CORRENTE = CC.COD_CONTA_CORRENTE
    LEFT JOIN NBS.TIPO_PAGAMENTO TP ON O.COD_EMPRESA = TP.COD_EMPRESA
                                                AND O.COD_TIPO_PAGAMENTO = TP.COD_TIPO_PAGAMENTO
    WHERE O.cod_operacao in (-1, --Peças Balcão
                            --2, --Serviços
                            --3, --Serviços Garantias
                            4, --Máquinas
                            9, --Máquinas
                            --7, --Peças Devolução
                            --13,
                            --19, --Peças Atacado
                            22, --Diversos
                            --26
                            60, --Devolução de Máquinas fins diversos
                            28 --Máquinas Devolução
                            )
)
SELECT TRUNC(CAB.EMISSAO) AS DATA_FATURAMENTO
        , CAB.CONTROLE
        , CAB.SERIE
        , ED.DESCRICAO AS DEPARTAMENTO
        , CAB.VENDEDOR
        , CAB.COD_CLIENTE_NOTA
        , CAB.COD_CLIENTE
        , CAB.COD_OPERACAO
        , CAB.TOTAL_NOTA AS VALOR_VENDIDO
        , CAB.VALOR_ICMS*(-1) AS VALOR_ICMS
        , CAB.ICMS_SUBSTITUIDO
        , CAB.ANO_MODELO
        , CAB.ANO_FABRICACAO
        , CAB.VALOR_PIS*(-1) AS VALOR_PIS
        , CAB.VALOR_COFINS*(-1) AS VALOR_COFINS
        , CAB.STATUS
        , VS.DESCRICAO AS DESCRSTATUS
        , CAB.END_CID_DESCRICAO
        , CAB.END_CID_UF
        , CAB.END_CID_DESCRICAO || '' || CAB.END_CID_UF AS C
--Dados da Fabrica
```

Data Transfer

Confirm

Check results

Export target  
Extraction settings  
Format settings  
Output  
Confirm

Objects

Source Container	Source	Target Container	Target
NBS_TESTE	WITH OPERACAO A...	C:\Users\jose.silva\...	VENDAS.csv

Source settings

Table settings:

- Open new connection(s): No
- Extract type: SEGMENTS
- Select row count: No
- Selected rows only: No
- Selected columns only: No

Target settings

File settings:

- Write to the single file: No
- Directory: C:\Users\jose.silva\Documents
- File name pattern: VENDAS
- On object data file name conflict: Autofix name
- On blob value file name conflict: Autofix name
- Encoding: UTF-8
- Timestamp pattern: yyyyMMddHHmm
- Insert BOM: No
- Compress: No
- Binaries: SKIP
- Encoding: RINARV

Save task    Back    Next >    Proceed    Cancel

Pesquisar 200 200+ 200 row(s) fetched - 6.487s (24ms fetch), on 2023-09-21 at 11:18:20

30°C Ensolado 11:20 POR PTB2 21/09/2023

## 2.2. Executando query e gerando o CSV GRUPOSEQP

The screenshot shows the DBeaver interface with a SQL Editor window containing a complex multi-table JOIN query. The results pane displays a table with two columns: COD\_NOTA\_MODELO and DESCRIAO\_NOTA\_MODELO. The table has 10 rows, with the first row highlighted.

COD_NOTA_MODELO	DESCRIAO_NOTA_MODELO
1	213
2	213
3	217
4	208
5	209
6	211
7	215
8	216
9	214
10	218

## 2.3. Executando query e gerando o CSV LOJAS

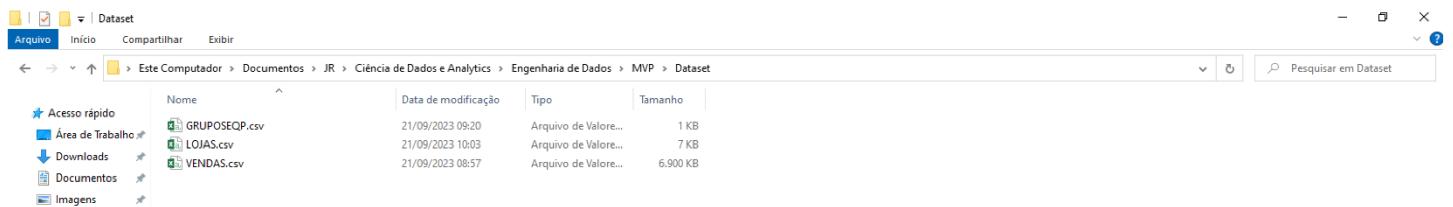
The screenshot shows the DBeaver interface with a SQL Editor window containing a query to select all columns from the EMPRESAS table ordered by COD\_EMPRESA. The results pane displays a table with several columns: COD\_EMPRESA, NOME, STATUS, INSCRICOAO\_ESTADUAL, LOCAL, CGC, CONTATO, INSCRICOAO\_MUNICIPAL, and FACHADA. The table has 7 rows, with the first row highlighted.

COD_EMPRESA	NOME	STATUS	INSCRICOAO_ESTADUAL	LOCAL	CGC	CONTATO	INSCRICOAO_MUNICIPAL	FACHADA
1		S		S				345
2		S		N				N. 425
3		S		N				694
4		S		N				0
5		S		N				2.257
6		S		N				3525
7		S		N				N. 425

## 3. Coleta dos Dados

### 3.1. Carregando os CSVs para o Azure

Foi criada uma conta de armazenamento no Azure e criado um container (mvp3container) para onde fizemos o upload do CSVs que estavam em minha máquina local conforme imagens abaixo. Esta operação consiste em colocarmos os CSVs no Datalake criado.



Segue abaixo a imagem mostrando os arquivos CSVs após upload para o Datalake mvp3container.

A screenshot of the Microsoft Azure Storage Container blade for the 'mvp3storage\_1695308110492' account. The container is named 'mvp3container'. The left sidebar shows options like 'Visão geral', 'Diagnosticar e resolver problemas', 'IAM (Controle de Acesso)', 'Configurações', 'Tokens de acesso compartilhado', 'Política de acesso', 'Propriedades', and 'Metadados'. The main area displays the uploaded CSV files: 'GRUPOSEQP.csv', 'LOJAS.csv', and 'VENDAS.csv'. Each file entry includes a checkbox, a preview icon, the name, last modified date (21/09/2023), access tier ('Principal (Inferidos)'), status ('Disponível'), blob type ('Blob de blocos'), size, and concession state ('Estado de concessão'). A toolbar at the top provides actions like 'Carregar', 'Alterar o nível de acesso', 'Atualizar', 'Excluir', 'Alterar a camada', 'Adquirir concessão', 'Interromper concessão', 'Exibir instantâneos', and 'Criar instantâneo'. A search bar at the bottom allows filtering by prefix.

## 3.2. Recursos criados

Foram criados alguns recursos na nuvem e um deles é o banco de dados SQL Server (mvp3sql) para onde faremos a ingestão dos dados após fazermos o ETL utilizando o recurso Data Factory (mvp3factory) que foi criado.

Serviços do Azure

Criar um recurso Todos os recursos Gerenciamento de custo +... Data factories Grupos de recursos Assinaturas Bancos de dados SQL Contas de armazenamento Centro de Início Rápido Mais serviços

Recursos

Recente Favorito

Nome	Tipo	Última visualização
mvp3factory	Data factory (V2)	16 minutos atrás
mvp3keyvault	Cofre de chaves	4 horas atrás
mvp3sql	SQL Server	4 horas atrás
MeuCatalogo	Conta do Microsoft Purview	5 horas atrás
MVP	Grupo de recursos	5 horas atrás
Azure subscription 1	Assinatura	6 horas atrás
mvp3sql	Banco de dados SQL	18 horas atrás
mvp3storage	Conta de armazenamento	5 dias atrás

Ver todos

Navegar

Assinaturas Grupos de recursos Todos os recursos Painel

Pesquisar

36°C Panc. de chuva POR PTB2 15:03 27/09/2023

## 4. Carga dos Dados

O ETL foi realizado utilizando Data Factory. Foi feito primeiramente a etapa de extração dos dados criando um dataset para cada arquivo csv que tinham como origem o storage account no container mvp3container (Datalake). Na segunda etapa foi feita a transformação através da criação de um dataflow para fazer as conversões e eliminação de atributos. Na terceira etapa foi feita a ingestão dos dados no banco de dados SQL Server mvp3sql.

### 4.1. Extração

Criei um dataset para cada arquivo csv para posteriormente incluirmos os mesmos como sources de um dataflow.

The screenshot shows the Microsoft Azure Data Factory interface. On the left, the 'Recursos do Data Factory' sidebar lists various components: Pipelines (pipeline1), Captura de Dados de Alterações..., Conjuntos de Dados (4 items: AzureSqlTable1, Grupos, Lojas, Vendas), and Fluxos de dados (dataflow1, Power Query). In the main workspace, a dataset named 'Grupos' is being created. It is identified as a 'DelimitedText' type with a 'CSV' file icon. The 'Conexão' tab is selected, showing the connection configuration: 'Serviço vinculado' set to 'AzureBlobStorage1', 'Caminho do arquivo' set to 'mvp3container / Diretório / GRUPOSEQP.csv', and other settings like 'Delimitador de colunas' (Comma ()), 'Delimitador de linha' (Default (\r\n, or \r\n)), 'Codificação' (Padrão(UTF-8)), 'Caractere de aspas' (Double quote (")), and 'Caractere de escape' (Backslash (\)). The 'Pesquisar' bar at the bottom has 'Lojas' selected. The status bar at the bottom right shows weather (37°C, Pred ensolarado), date (25/09/2023), and time (15:36).

Observem que no momento da criação do dataset informamos qual é o delimitador das colunas do csv para importarmos os atributos corretamente, mas todos são reconhecidos como string e posteriormente no processo de transformação vamos colocar o tipo correto e a formatação.

This screenshot is nearly identical to the previous one, showing the creation of a new dataset named 'Lojas'. The 'Conexão' tab is selected, and the configuration is identical to the 'Grupos' dataset: 'Serviço vinculado' set to 'AzureBlobStorage1', 'Caminho do arquivo' set to 'mvp3container / Diretório / LOJAS.csv', and other settings like 'Delimitador de colunas' (Comma ()), 'Delimitador de linha' (Default (\r\n, or \r\n)), 'Codificação' (Padrão(UTF-8)), 'Caractere de aspas' (Double quote (")), and 'Caractere de escape' (Backslash (\)). The 'Pesquisar' bar at the bottom has 'Lojas' selected. The status bar at the bottom right shows weather (37°C, Pred ensolarado), date (25/09/2023), and time (15:38).

Aqui foi criado um dataset para o csv Vendas e observem que na visualização dos dados o atributo DATA\_FATURAMENTO veio como string e no processo de transformação vamos colocar como tipo DATE.

	DATA_FATURAMENTO	CONTROLE	SERIE	DEPARTAMENTO	VENDEDOR	COD_CLIENTE_NOTA	COD
1	2013-08-28T03:00:00Z	72254	1	Administracao	MTAVEIRA		1037
2	2013-08-28T03:00:00Z	72250	1	Administracao	MTAVEIRA		1037
3	2013-08-28T03:00:00Z	72251	1	Administracao	MTAVEIRA		1037
4	2017-12-29T02:00:00Z	59336	1	Administracao	MMACHADO		1696
5	2013-10-04T03:00:00Z	74418	1	Administracao	MTAVEIRA		8179

## 4.2. Transformação

Criei um dataflow e nele foram inseridos os datasets criados anteriormente. Neste passo foi definido os tipos dos atributos de cada dataset, também o formato. Nesse momento poderíamos renomear os atributos.

Nome da coluna de saída *	Coluna do source1	Nomear como
DATA_FATURAMENTO	DATA_FATURAMENTO	
CONTROLE	CONTROLE	

Nesta etapa de definição das fontes de dados do dataflow foram feitas as atribuições dos tipos corretos dos atributos e as formatações corretas dos atributos. Na imagem abaixo e relatado anteriormente, o atributo DATA\_FATURAMENTO do dataset VENDAS foi definido agora como tipo DATE com formato yyyy-MM-dd. Atributos que representam moeda foram definidos como LONG e alguns somente numéricos como INTEGER.

Foram renomeados alguns atributos para ficarem com a mesma identificação entre tabelas fato e tabela dimensão (Exemplo "GRUPO" no dataset Vendas e no dataset Grupos).

A Microsoft anuncia recentemente a versão prévia pública do Microsoft Fabric, uma maneira nova e interessante de criar a análise de dados em nuvem. Clique [aqui](#) para começar a usar o Fabric Data Factory!

Experiência de versão prévia Desligado

Pesquisar alocador e documentação

joserobertofox@hotmail.com DIRETÓRIO Padrão

37°C Pred ensolarado 16:04 POR PTB 25/09/2023

Depois dos tratamentos dos atributos incluímos no dataflow 2 JOINs. Fizemos isto para deixarmos apenas uma tabela desnormalizada na qual foi incluído o atributo nome da loja pelo primeiro join e o atributo nome do grupo no segundo join, servirão para concluirmos o nosso objetivo principal. Com isso preparamos os nossos dados para fazer a ingestão no banco de dados SQL Server (mvp3sql). Foi feito nas duas imagens seguintes o primeiro JOIN1 entre o dataset VENDAS (select1) e o dataset LOJAS (select2) sendo utilizada o atributo COD\_EMPRESA dos dois datasets para incluirmos no dataset VENDAS a coluna que contém o nome da loja.

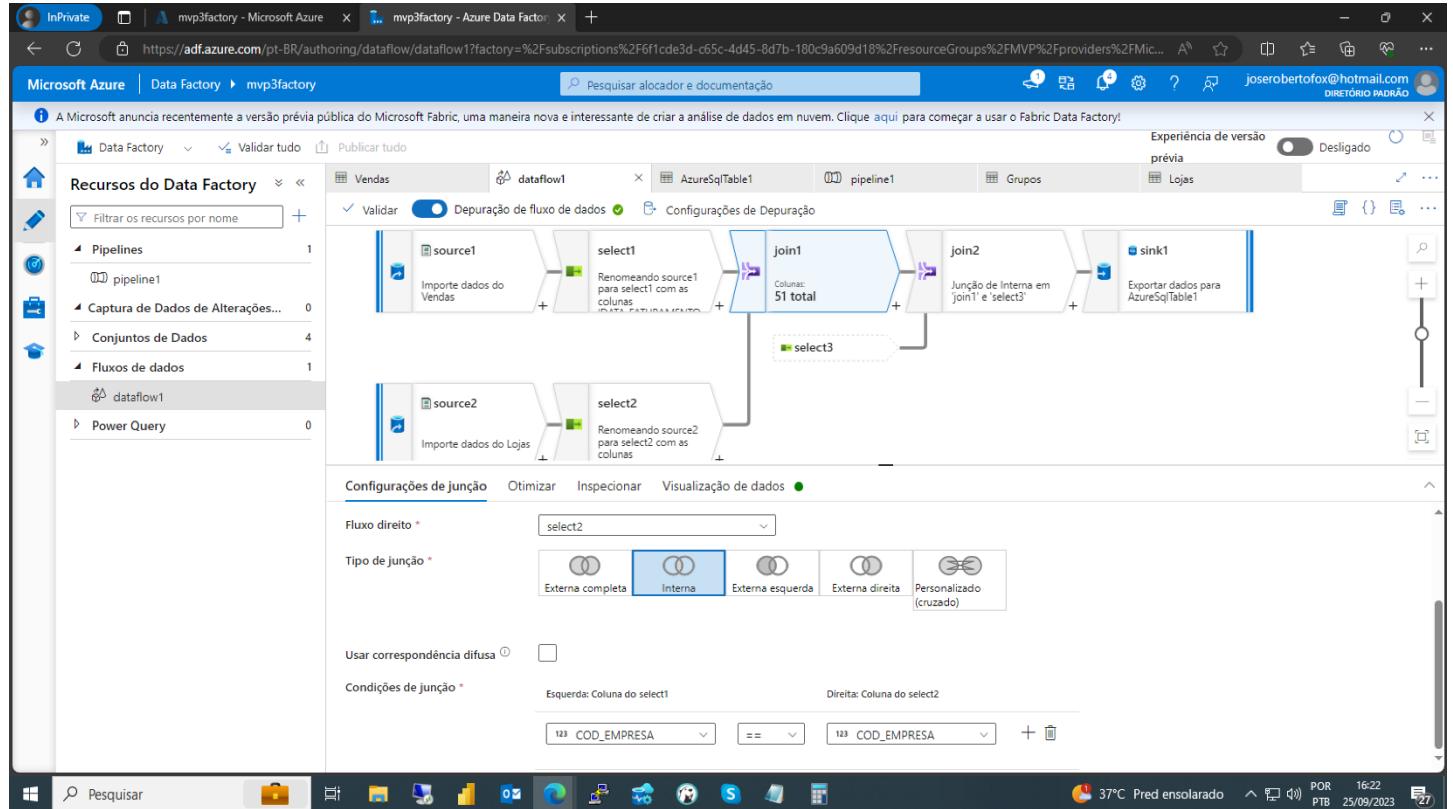
A Microsoft anuncia recentemente a versão prévia pública do Microsoft Fabric, uma maneira nova e interessante de criar a análise de dados em nuvem. Clique [aqui](#) para começar a usar o Fabric Data Factory!

Experiência de versão prévia Desligado

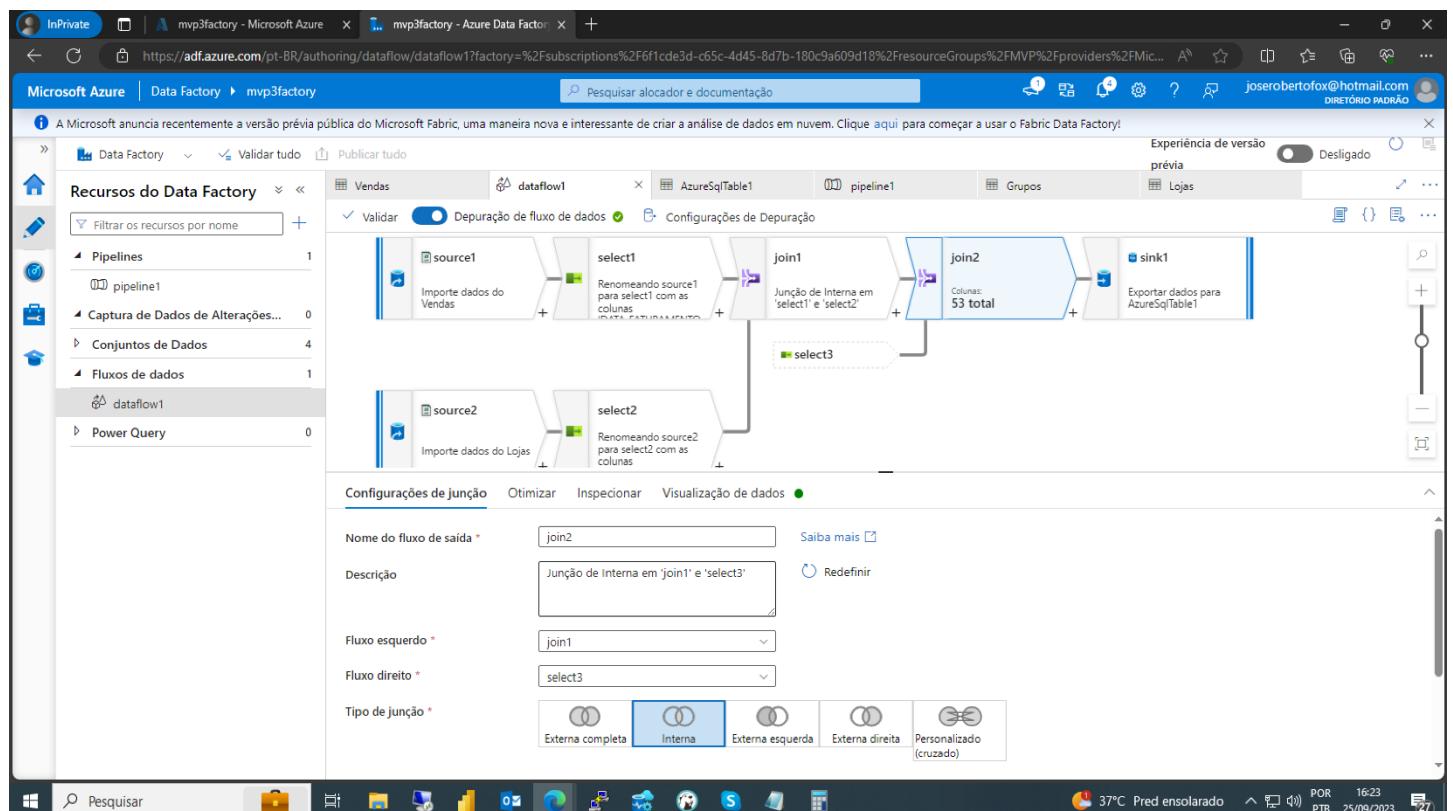
Pesquisar alocador e documentação

joserobertofox@hotmail.com DIRETÓRIO Padrão

37°C Pred ensolarado 16:21 POR PTB 25/09/2023



Foi feito nas duas imagens a seguir o segundo JOIN2 entre o JOIN1 e o dataset GRUPOS (select3) sendo utilizado o atributo GRUPO dos dois datasets para incluirmos no dataset VENDAS a coluna que contém o nome do grupo do equipamento.



## 4.3. Carga dos Dados

Nesta etapa criamos no dataflow a exportação dos dados SINK1 para o dataset AzureSqlTable1 e vinculamos a ela o nosso banco de dados mvp3sql criado para receber nossos dados.

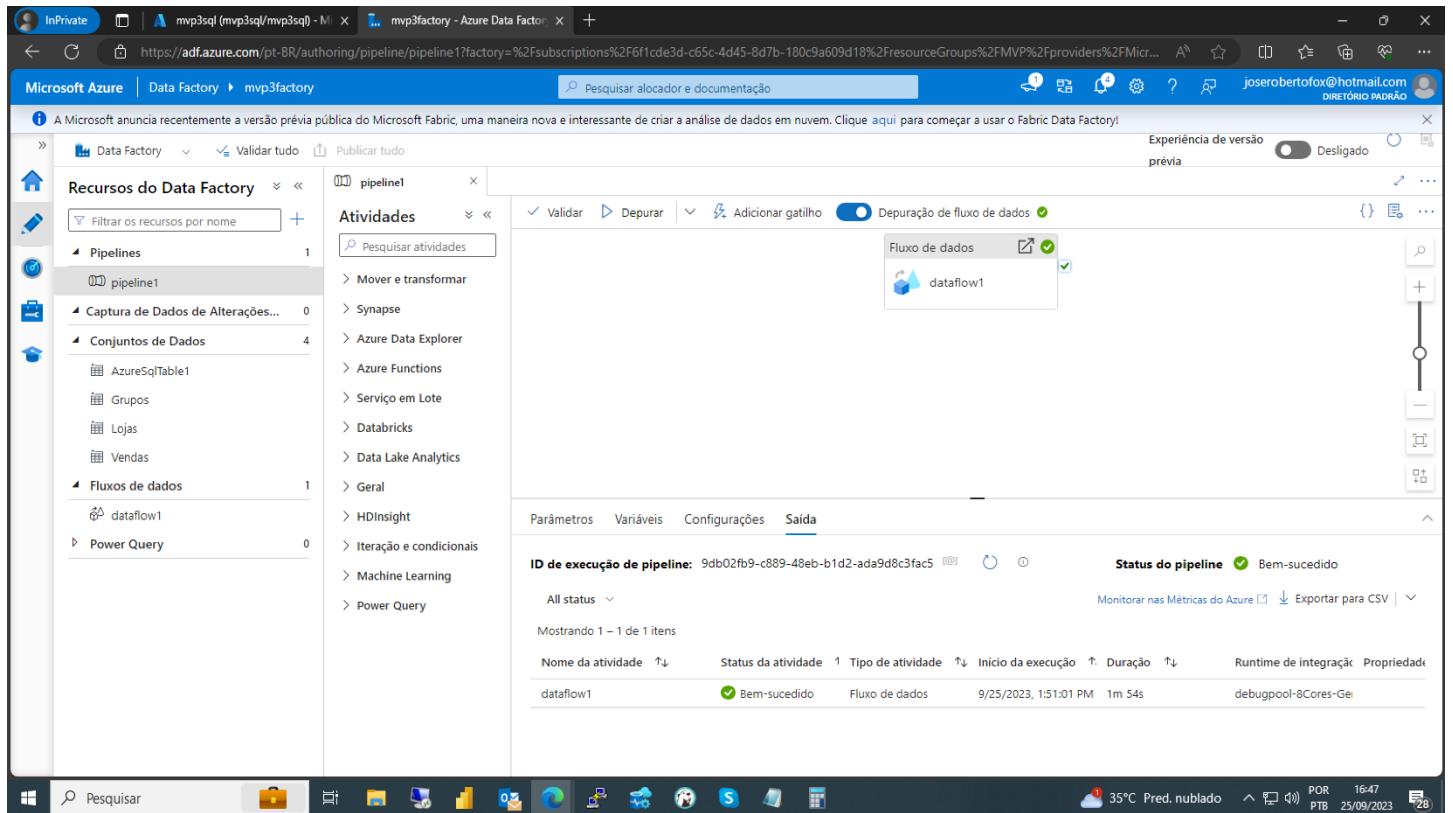
Passamos as informações de nome do domínio (mvp3sql.database.windows.net), nome do banco de dados (mvp3sql) e usuário e senha que criamos no início do projeto quando instalamos o recurso Banco de Dados SQL.

The screenshot shows the 'Edit linked service' dialog for 'AzureSqlTable1'. The left sidebar lists resources: Pipelines (1), Captura de Dados de Alterações... (0), Conjuntos de Dados (4), Grupos, Lojas, Vendas, and Fluxos de dados (1). The main area shows a connection to 'Banco de Dados SQL do Azure' with 'AzureSqlTable1'. The 'Conexão' tab is selected, showing 'Serviço vinculado' set to 'AzureSqlDatabase1' and 'Tabela' set to 'dbo.vendas'. On the right, the 'Editar serviço vinculado' panel is open, showing the connection details: Nome (AzureSqlDatabase1), Conectar-se via runtime de integração (AutoResolveIntegrationRuntime), Cadeia de conexão (Azure Key Vault), Método de seleção de conta (From Azure subscription), Nome de domínio totalmente qualificado (mvp3sql.database.windows.net), Nome do Banco de Dados (mvp3sql), Tipo de Autenticação (SQL authentication), Nome do usuário (mvp3admin), and Senha (Azure Key Vault). Buttons for 'Aplicar' and 'Cancelar' are at the bottom.

The screenshot shows the 'Visualizar dados' (View data) dialog for 'AzureSqlTable1'. The left sidebar is identical to the previous screenshot. The main area shows the connection to 'Banco de Dados SQL do Azure' with 'AzureSqlTable1'. The 'Conexão' tab is selected, showing 'Serviço vinculado' set to 'AzureSqlDatabase1' and 'Tabela' set to 'dbo.vendas'. On the right, the 'Visualizar dados' panel displays the data from the 'dbo.vendas' table in a grid format. The columns are: #, DATA\_FATURAMENTO, CONTROLE, SERIE, DEPARTAMENTO, VENDEDOR, COD\_CLIENTE\_NOTA, and COD. The data rows are:

#	DATA_FATURAMENTO	CONTROLE	SERIE	DEPARTAMENTO	VENDEDOR	COD_CLIENTE_NOTA	COD
1	08/28/2013	72254	1	Administracao	MTAVEIRA		1037
2	08/28/2013	72250	1	Administracao	MTAVEIRA		1037
3	08/28/2013	72251	1	Administracao	MTAVEIRA		1037
4	12/29/2017	59336	1	Administracao	MMACHADO		1696
5	10/04/2013	74418	1	Administracao	MTAVEIRA		8179

Validamos, depuramos e publicamos o nosso Dataflow e incluímos ele no nosso Pipeline. Validamos e depuramos também o nosso Pipeline. Agora nosso banco de dados está populado e partiremos agora para a ferramenta Azure Data Studio para executarmos algumas queries e testarmos e analisarmos o nosso Data Warehouse.

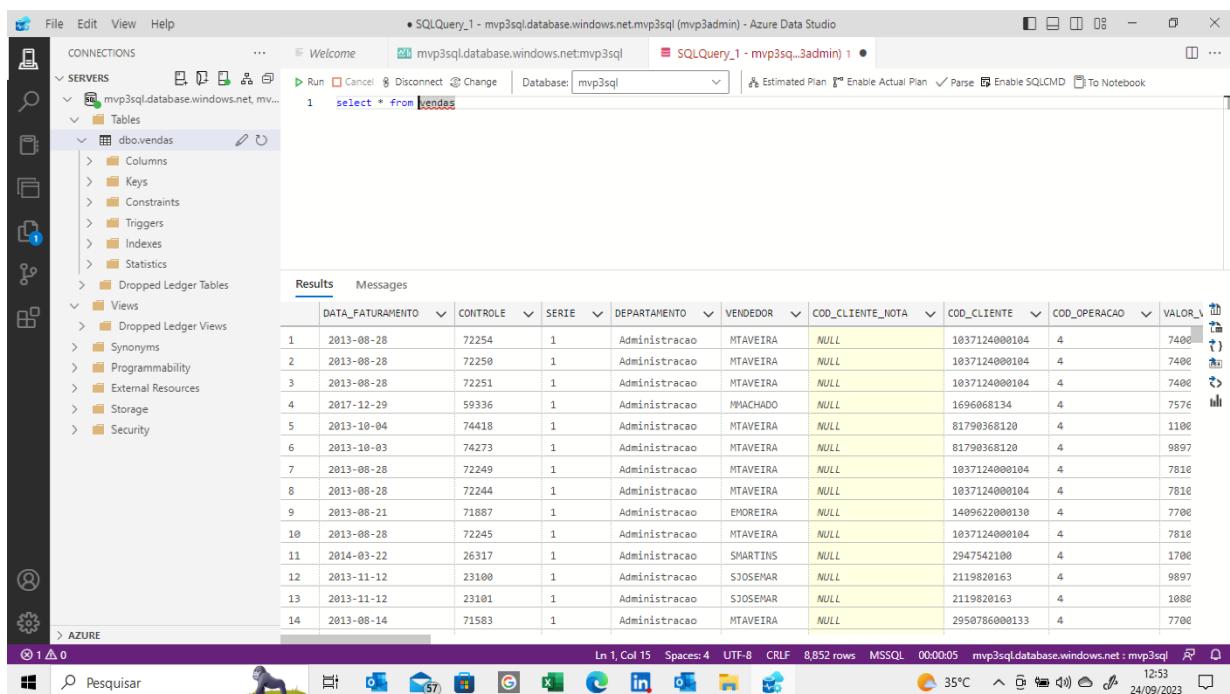


The screenshot shows the Microsoft Azure Data Factory interface. On the left, the navigation pane shows 'Recursos do Data Factory' with 'Pipelines' expanded, showing 'pipeline1'. Under 'Fluxos de dados', there is a 'dataflow1' item. The main area displays the 'Atividades' (Activities) tab for pipeline1. A search bar at the top of this section says 'Pesquisar atividades'. Below it is a list of activity types: Mover e transformar, Synapse, Azure Data Explorer, Azure Functions, Serviço em Lote, Databricks, Data Lake Analytics, Geral, HDInsight, Iteração e condicionais, Machine Learning, and Power Query. To the right, a 'Fluxo de dados' (Data Flow) card for 'dataflow1' is shown with a green checkmark. Below the activities list, tabs for 'Parâmetros', 'Variáveis', 'Configurações', and 'Saída' are visible. The 'Saída' tab is selected. It shows the 'ID de execução de pipeline' as 9db02fb9-c889-48eb-b1d2-ada9d8c3fac5 and the 'Status do pipeline' as 'Bem-sucedido'. A table below lists the execution details: Nome da atividade (dataflow1), Status da atividade (Bem-sucedido), Tipo de atividade (Fluxo de dados), Início da execução (9/25/2023, 15:01:01 PM), Duração (1m 54s), and Runtime de integração (debugpool-8Cores-Ge). At the bottom of the interface, there are various Azure service icons and a status bar showing weather (35°C, Pred. nublado), date (25/09/2023), and time (16:47).

#### 4.4. Utilizando o Azure Data Studio para testar o database

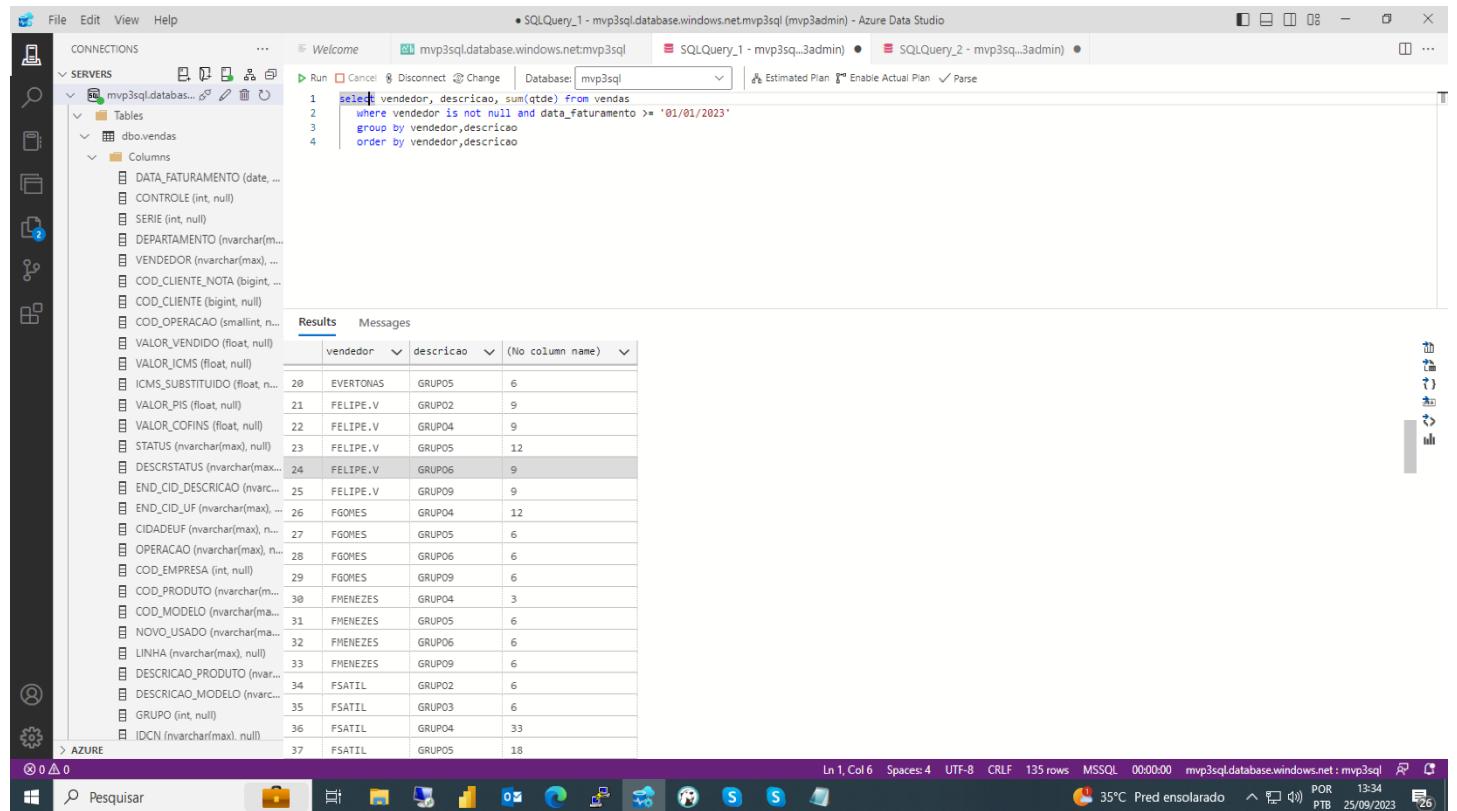
Criei a conexão do banco de dados com Azure Data Studio e executei algumas queries fazendo referência aos objetivos a serem atendidos. Estas queries servirão para compararmos as quantidades com os resultados obtidos no dashboard que será criado.

SELECT 1



The screenshot shows the Azure Data Studio interface. The left sidebar shows 'CONNECTIONS' with a connection to 'mvp3sql.database.windows.net:mvp3sql'. Under 'TABLES', there is a 'dbo.vendas' table. The main area displays the results of a query: 'select \* from vendas'. The results grid has columns: DATA\_FATURAMENTO, CONTROLE, SERIE, DEPARTAMENTO, VENDEDOR, COD\_CLIENTE\_NOTA, COD\_CLIENTE, COD\_OPERACAO, and VALOR\_. The data shows various sales entries with dates ranging from 2013-08-28 to 2014-03-22, and values ranging from 7400 to 9897. At the bottom of the interface, there is a toolbar with various icons and a status bar showing 'Ln 1, Col 15, Spaces: 4, UTF-8, CRLF, 8.852 rows, MSSQL, 00:00:05, mvp3sql.database.windows.net:mvp3sql' and system information like temperature (35°C), date (24/09/2023), and time (12:53).

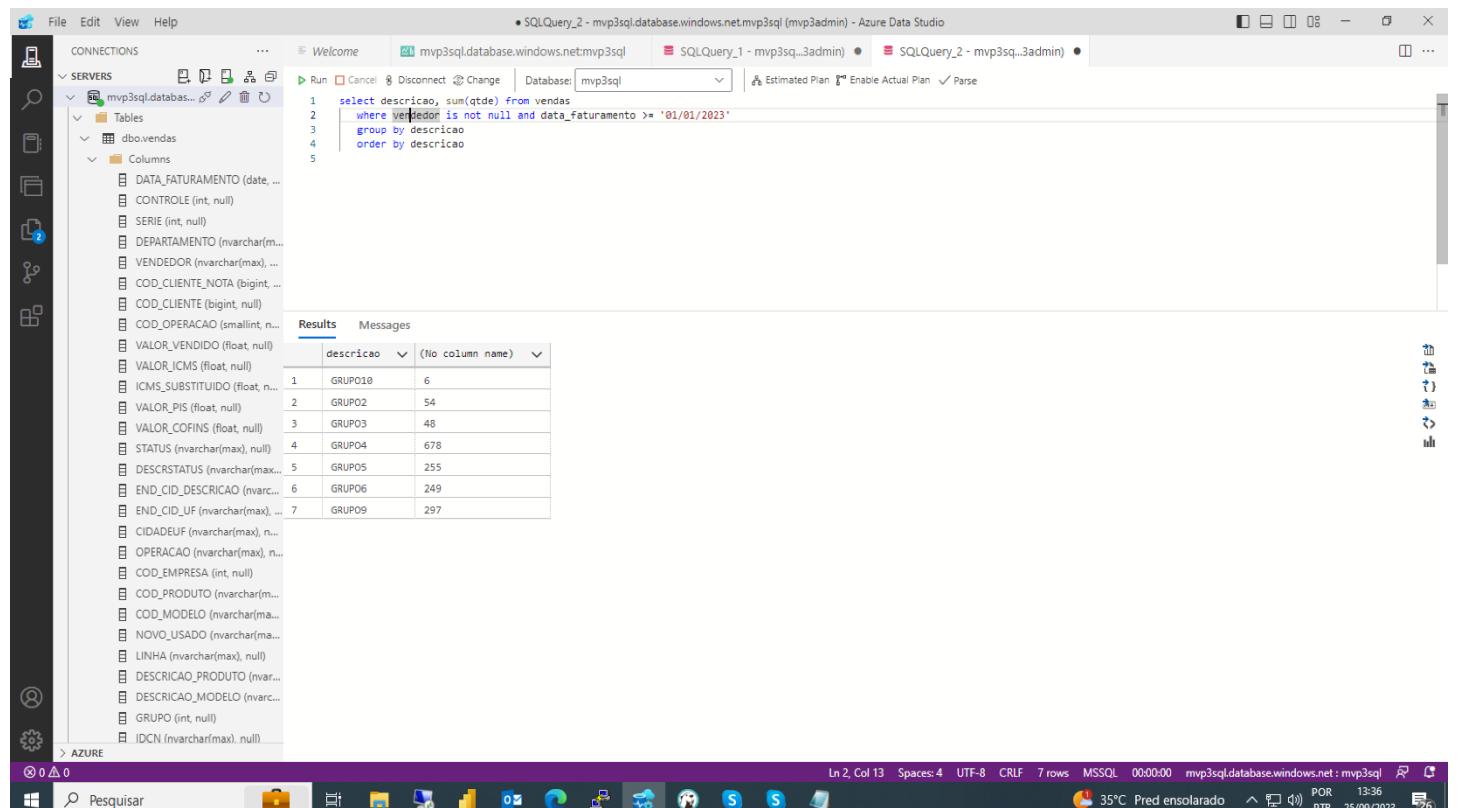
## SELECT 2 - Select com agrupamento por vendedor e descrição do grupo de equipamento.



Azure Data Studio interface showing a query results grid. The results show sales grouped by vendor and product category. The vendor column has 37 distinct values, and the product category column has 9 distinct values. The total sales per vendor range from 6 to 678.

vendedor	descricao	(No column name)
20	EVERTONAS	GRUPOS
21	FELIPE.V	GRUPO2
22	FELIPE.V	GRUPO4
23	FELIPE.V	GRUPOS
24	FELIPE.V	GRUPO6
25	FELIPE.V	GRUPO9
26	FGOMES	GRUPO4
27	FGOMES	GRUPOS
28	FGOMES	GRUPO6
29	FGOMES	GRUPO9
30	FMENEZES	GRUPO4
31	FMENEZES	GRUPOS
32	FMENEZES	GRUPO6
33	FMENEZES	GRUPO9
34	FSATIL	GRUPO2
35	FSATIL	GRUPO3
36	FSATIL	GRUPO4
37	FSATIL	GRUPOS

## SELECT 3 - Select com totais vendidos por descrição do grupo de equipamentos.



Azure Data Studio interface showing a query results grid. The results show sales grouped by product category. The product category column has 7 distinct values. The total sales per category range from 6 to 678.

descricao	(No column name)
1	GRUPO10
2	GRUPO2
3	GRUPO3
4	GRUPO4
5	GRUPOS
6	GRUPO6
7	GRUPO9

## SELECT 4 - Select com totais vendidos por código do grupo de equipamentos.

The screenshot shows the Azure Data Studio interface. In the top right, there are tabs for 'Welcome' and 'mvp3sql.database.windows.net|mvp3sql'. Below them, a query window displays the following SQL code:

```
1 select grupo, sum(qtde) from vendas
2 where vendedor is not null and data_faturamento >= '01/01/2023'
3 group by grupo
4 order by grupo
```

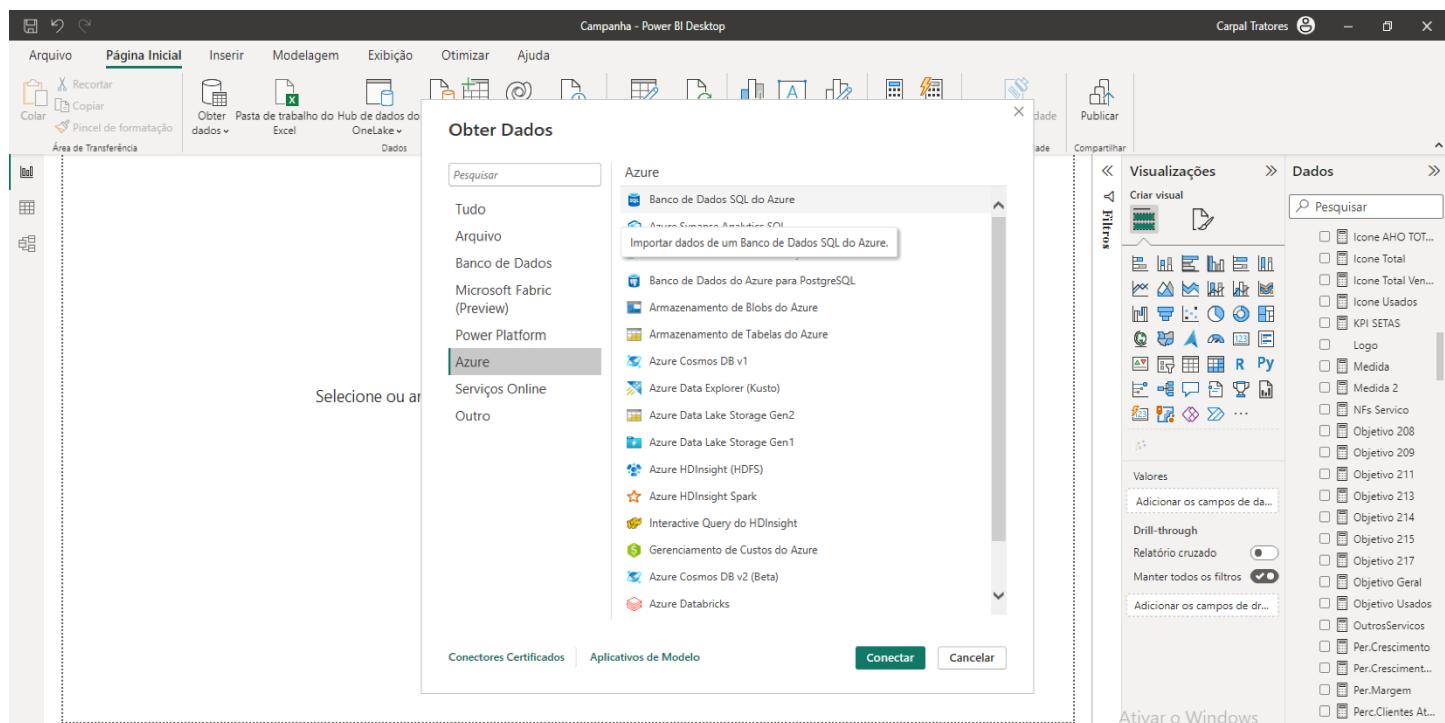
The results pane shows a table with two columns: 'grupo' and '(No column name)'. The data is as follows:

grupo	(No column name)
1	208
2	209
3	211
4	213
5	214
6	217
7	218

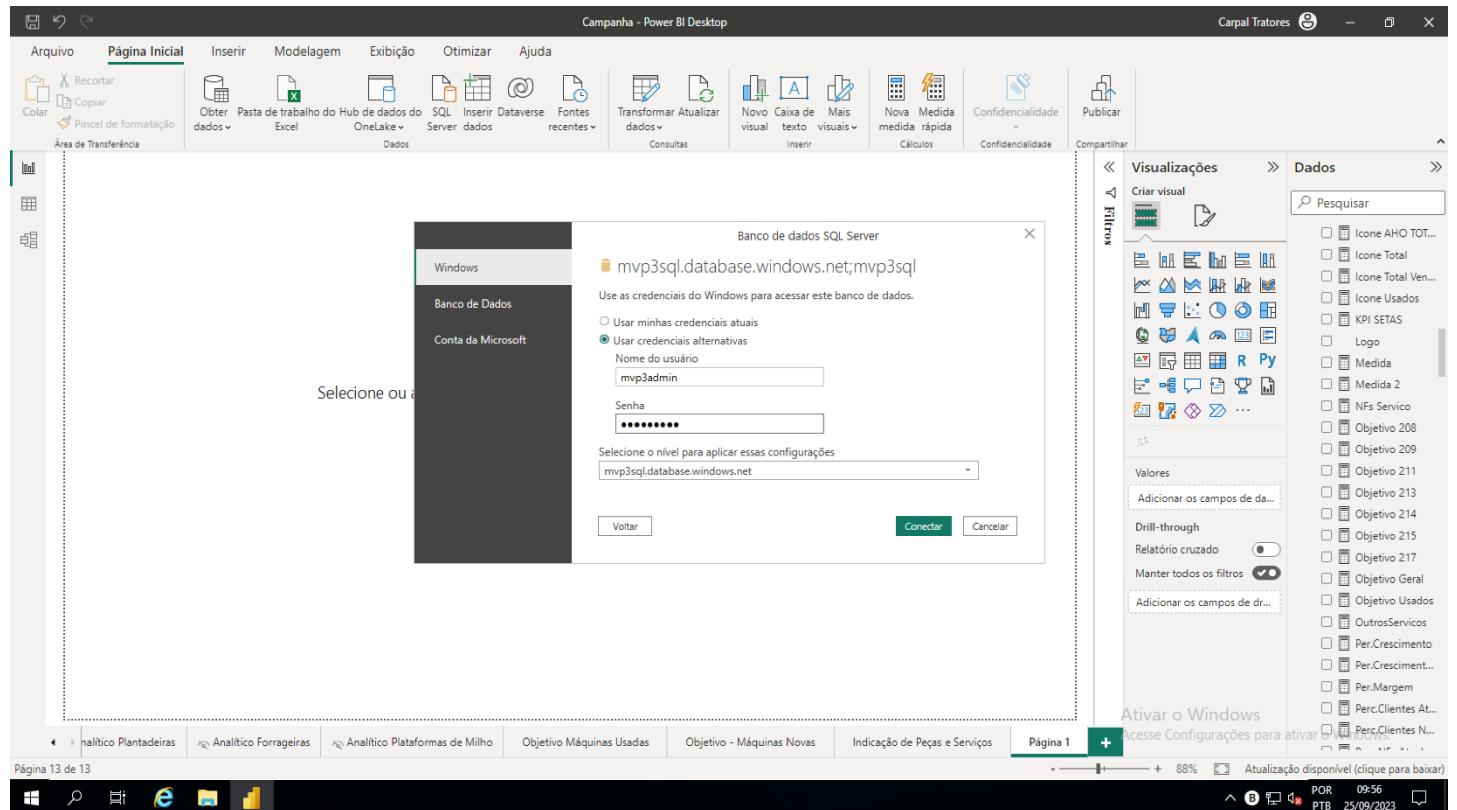
At the bottom of the interface, there is a toolbar with various icons and status information like 'Ln 1, Col 13', 'Spaces: 4', 'UTF-8', 'CRLF', '7 rows', 'MSSQL', '00:00:00', 'mvp3sql.database.windows.net : mvp3sql', '35°C Pred ensolarado', 'POR PTB 25/09/2023', and a battery icon.

## 4.5. Carga do database (mvp3sql) na Ferramenta de BI (POWER BI)

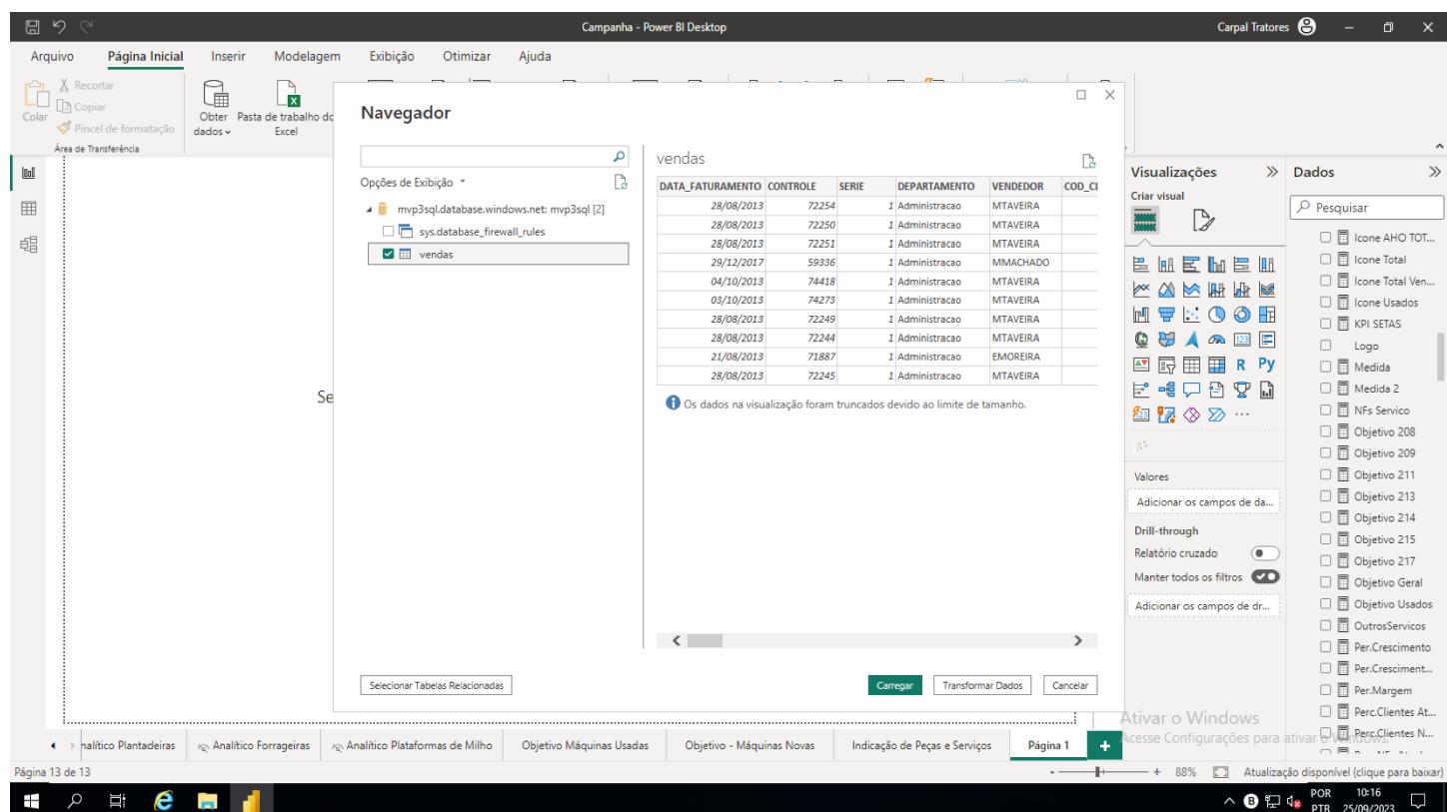
Vou utilizar o POWER BI para fazer o dashboard para atender o objetivo, desta forma vamos criar uma conexão com o banco de dados na nuvem Azure para obtermos os dados para o POWER BI. Demonstramos abaixo o processo de obter no POWER BI Desktop e selecionando a fonte BANCO DE DADOS SQL DO AZURE.



Fornecemos as credenciais cadastradas no AZURE e conectamos.



O POWER BI Desktop localiza o database e a tabela VENDAS é selecionada sendo carregada para desenvolvemos o dashboard.



## 5. Modelagem

Foi utilizado o Microsoft Purview para fazer a modelagem da tabela VENDAS. Fizemos a conexão do Purview com o nosso database, criamos um scan e localizamos nossa Azure Sql Table VENDAS para criarmos o catálogo dos atributos.

The screenshot shows the Microsoft Purview Data sources interface for the 'MeuCatalogo' catalog. On the left, there's a sidebar with various navigation options like Data sources, Collections, Monitoring, Metamodel, etc. The main area shows a 'Data sources' list with one item: 'Mvp3' (Azure SQL Database). To the right, there's a 'Scope your scan' panel with a 'Refresh' button and a note about selecting assets under a parent. A search bar and a tree view for selecting specific assets are also present. At the bottom, there are 'Continue' and 'Back' buttons.

This screenshot shows the 'Edit data source (Azure SQL Database)' configuration dialog for the 'Mvp3' source. It includes fields for 'Data source name' (set to 'Mvp3'), 'Azure subscription' (set to 'Azure subscription 1 (6f1cde3d-c65c-4d45-8d7b-180c9a609d18)'), 'Server name' (set to 'mvp3sql'), 'Endpoint' (set to 'mvp3sql.database.windows.net'), and 'Collection' (set to '[Root] MeuCatalogo'). A note indicates that all items in the data source will belong to this collection. Under 'Data policy enforcement (previously Data use management)', a switch is set to 'off'. At the bottom, there are 'Apply' and 'Cancel' buttons.

Abaixo mostramos a linhagem da tabela vendas que originou do nosso pipeline.

The screenshot shows the Microsoft Purview Data Catalog interface. On the left, there's a navigation pane with icons for InPrivate, Microsoft Azure, and MeuCatalogo. Below that, it says 'Browse assets' and 'By collection'. Under 'MeuCatalogo', there are sections for 'Sub collection(s)' and 'Related'. A sidebar on the left lists 'Narrow results by' categories: Data (Dashboard, Data pipeline, Data share, Database, File, Folder, Report, Stored procedure, Table) and Business (Application service, Business process, Data domain, Glossary terms, Product). The main content area shows a list of assets: 'vendas' (Azure SQL Table), 'dbo' (Azure SQL Schema), 'mvp3sql' (Azure SQL Database), and 'mvp3sql.database.windows.net' (Azure SQL Server). Each asset has a description, fully qualified name, and last updated time. At the bottom, there are buttons for 'Previous', 'Page 1 of 1', and 'Next'.

Catalogando os atributos da tabela VENDAS.

The screenshot shows the 'Edit "vendas"' page in the Microsoft Purview Data Catalog. The top navigation bar includes 'InPrivate', 'mvp3keyvault - Microsoft Azure', 'Edit vendas - MeuCatalogo', and a search bar. The left sidebar shows 'Data catalog > Browse assets > vendas'. The main content area is titled 'Edit "vendas"'. It has tabs for 'Overview', 'Schema' (which is selected), 'Lineage', and 'Contacts'. Below the tabs is a 'Filter by name' input field. The table below shows 13 columns: DATA\_FATURAMENTO (date, Data do Faturamento), CONTROLE (int, Número da nota fiscal de venda), SERIE (int, Série da nota fiscal de venda), DEPARTAMENTO (nvarchar, Código do departamento da loja), VENDEDOR (nvarchar, Nome do vendedor que efetuou a venda), COD\_CLIENTE (bigint, Código do cliente que houve o faturamento), COD\_OPERACAO (smallint, Código da operação (Venda/Devolução)), VALOR\_VENDIDO (float, Valor faturado), VALOR\_ICMS (float, Valor do icms da nota fiscal), and ICMS\_SUBSTITUIDO (float, Valor do icms de substituição tributária). At the bottom, there are 'Save' and 'Cancel' buttons.

Catalogando os atributos da tabela VENDAS. Como exemplo cito o atributo categórico STATUS com seus possíveis valores.

The screenshot shows the Microsoft Purview Data Catalog interface. The URL is https://web.purview.azure.com/resource/meucatalogo/main/catalog/entity/edit?guid=089e8d9e-2d17-4656-a623-b7f6f60000&feature.tenant=a919c871-3655-40b2-bee9-ea... . The page title is "Edit vendas - MeuCatalogo". The left sidebar shows "Data catalog > Browse assets > vendas > Edit "vendas"". The main area is titled "Edit "vendas"" and has tabs for "Overview", "Schema" (selected), "Lineage", and "Contacts". A search bar at the top says "Search catalog". Below the tabs, there is a "Filter by name" dropdown. The table structure for the "vendas" table is displayed with columns: COD\_OPERACAO, VALOR\_VENDIDO, VALOR\_ICMS, ICMS\_SUBSTITUIDO, VALOR\_PIS, VALOR\_COFINS, and STATUS. Each column has a "Select..." dropdown, a type dropdown (e.g., smallint, float, nvarchar), and a description. The STATUS column is described as "Identifica a situação da nota fiscal." with possible values: 0 - Ativa, 1 - Cancelada, 2 - Devolução Parcial, 3 - Devolução Total. At the bottom are "Save" and "Cancel" buttons. The taskbar at the bottom shows various icons and the date/time as 13:50 27/09/2023.

Mostramos na imagem abaixo que o banco de dados que contém a tabela VENDAS faz parte do catálogo de dados criado no Microsoft Purview.

The screenshot shows the Microsoft Purview Data Catalog interface. The URL is https://web.purview.azure.com/resource/meucatalogo/main/datasource/registeredSources?feature.tenant=a919c871-3655-40b2-bee9-ea3fa86a0d7d . The page title is "Data sources - MeuCatalogo". The left sidebar shows "Data sources" and "MeuCatalogo". The main area shows "Showing 1 collection, 1 data source". A tree view shows "MeuCatalogo" (The root collection) expanded to show "Mvp3" (Azure SQL Database). The "View details" button is visible for both. The taskbar at the bottom shows various icons and the date/time as 14:08 27/09/2023.

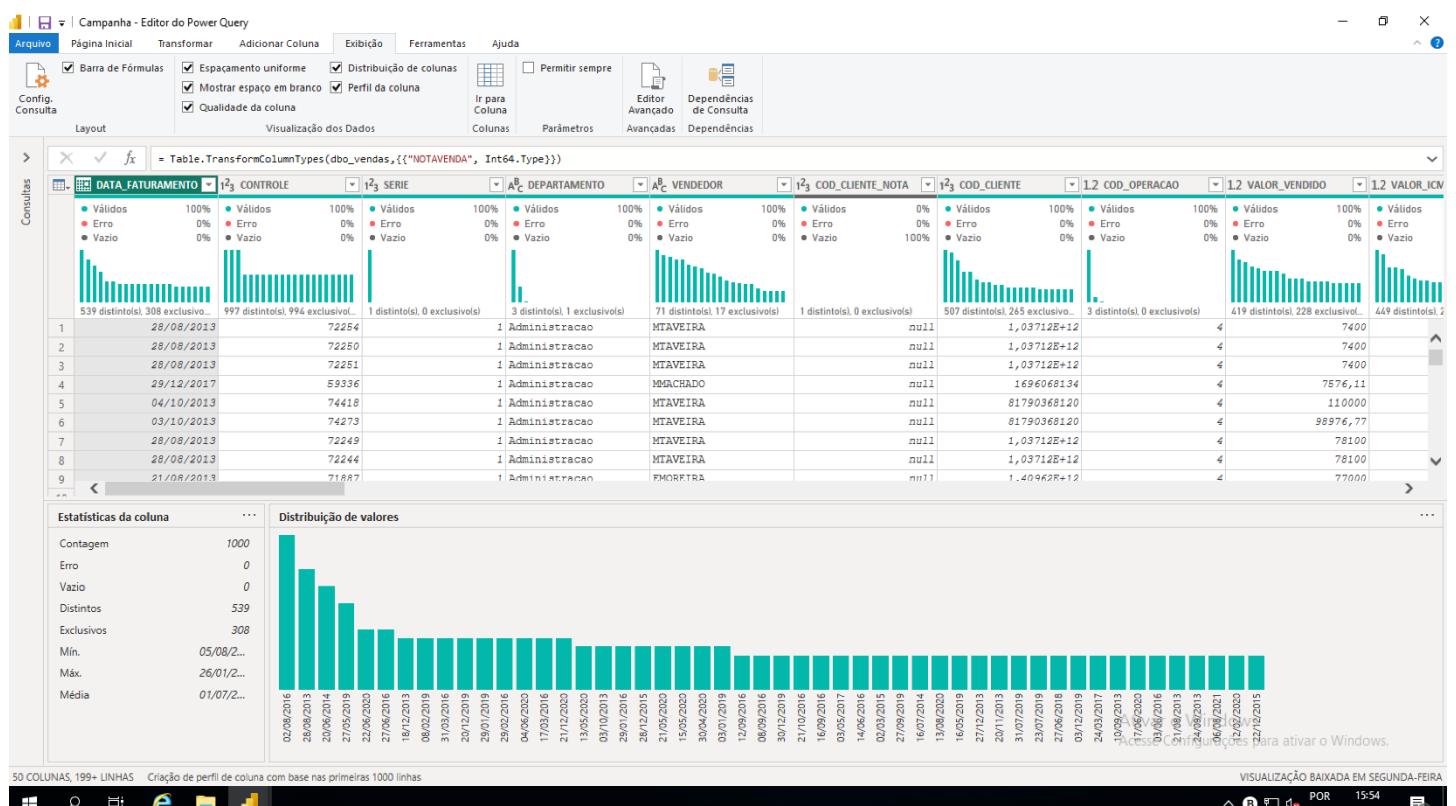
# 6. Análise dos Dados

## 6.1. Qualidade dos dados

Analisei os atributos do meu conjunto de dados que será utilizado para resolver o objetivo e verifiquei que não há problemas pois ele já veio com um tratamento. Não há erros nos atributos e os valores faltantes aparecem apenas em atributos sem restrição de integridade. Como minha tabela VENDAS contém 50 atributos, mostrei abaixo algumas análises nos atributos que são base para construir meu dashboard para atender o objetivo.

Fiz a análise utilizando o Power Query dentro do Power BI.

Análise do atributo DATA\_FATURAMENTO que utilizaremos para fazer a segmentação por data de faturamento. Verificamos que os valores estão 100% válidos, nenhum erro e nenhum registro com valor deste atributo está vazio e nem nulo.



Análise do atributo VENDEDOR que vamos utilizar para fazer o ranking e o grid, sendo muito importante para atendermos o nosso objetivo. Verificamos que os valores estão 100% válidos, nenhum erro e nenhum registro com valor deste atributo está vazio e nem nulo.

The screenshot shows the Power Query Editor interface with the 'Exibição' tab selected. The main area displays the 'VENDEDOR' column from a table named 'DATA\_FATURAMENTO'. Below the table, there are two sections: 'Estatísticas da coluna' and 'Distribuição de valores'. The 'Distribuição de valores' section contains a horizontal bar chart showing the count of distinct values for each category: Válidos (1000), Erro (0), and Vazio (0). The chart also includes a legend for 'Válidos' (green), 'Erro' (red), and 'Vazio' (black).

Categoria	Quantidade
Válidos	1000
Erro	0
Vazio	0

O Atributo QTDE abaixo contém somente o valor "1" porque existe uma particularidade em nossas vendas de equipamentos que permitem somente 1 equipamento por NF, devido a algumas exigências das regras de negócio (business rules).

The screenshot shows the Power Query Editor interface with the 'Exibição' tab selected. The main area displays the 'QTDE' column from a table named 'NOTAVENDA'. Below the table, there are two sections: 'Estatísticas da coluna' and 'Distribuição de valores'. The 'Distribuição de valores' section contains a horizontal bar chart showing the count of distinct values for each category: Válidos (1000), Erro (0), and Vazio (0). The chart also includes a legend for 'Válidos' (green), 'Erro' (red), and 'Vazio' (black).

Categoria	Quantidade
Válidos	1000
Erro	0
Vazio	0

Análise do atributo NOME que identifica a LOJA, vamos utilizar para fazer a segmentação de dados por loja. Verificamos que os valores estão 100% válidos, nenhum erro e nenhum registro com valor deste atributo está vazio e nem nulo.

Consultas

Campanha - Editor do Power Query

Arquivo Página Inicial Transformar Adicionar Coluna Exibição Ferramentas Ajuda

Config. Consulta Layout Visualização dos Dados Consultas Colunas Parâmetros Avançadas Dependências de Consulta Dependências

**NOTADEVOLUCAO**

**QUIDE**

**SOMAQUANT**

**TICKET**

**DRILL**

**REFAT**

**INDICOU**

**NOME**

**DESCRIAO**

100% Válidos 6% Válidos 100% Válidos 100% Válidos 100% Válidos 100% Válidos < 1% Válidos 100% Válidos 100% Válidos  
0% Erro  
0% Vazio 94% Vazio 0% Vazio 0% Vazio 0% Vazio 0% Vazio 99% Vazio 0% Vazio 0% Vazio

exclusivo(s) 57 distinto(s), 56 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 2 distinto(s), 1 exclusivo(s) 5 distinto(s), 0 exclusivo(s) 9 distinto(s), 0 exclusivo(s)

11 26317 null1 1 S S S S N null LOJA4 GRUPOS  
12 23100 null1 1 S S S S N null LOJA4 GRUPO4  
13 23101 null1 1 S S S S N null LOJA4 GRUPO4  
14 71563 null1 1 S S S S N null LOJA2 GRUPOS  
15 104743 null1 1 S S S S N null LOJA2 GRUPOS  
16 21523 null1 1 S S S S N null LOJA4 GRUPOS  
17 21529 null1 1 S S S S N null LOJA4 GRUPOS  
18 21530 null1 1 S S S S N null LOJA4 GRUPOS  
19 22403 null1 1 S S S S N null LOJA4 GRUPOS

20

Estatísticas da coluna ...

Distribuição de valores

Contagem 1000

Erro 0

Vazio 0

Distintos 5

Exclusivos 0

Cadeia de caracteres vazia 0

Min. LOJA2

Máx. LOJA6

Ativar o Windows  
Acesse Configurações para ativar o Windows.

50 COLUNAS, 199+ LINHAS Criação de perfil de coluna com base nas primeiras 1000 linhas

VISUALIZAÇÃO BAIKADA EM SEGUNDA-FEIRA

POR PTB 26/09/2023

Análise do atributo DESCRIAO que identifica a descrição do grupo de equipamentos. Vamos utilizar este atributo para fazer o grid solicitado no objetivo. Verificamos que os valores estão 100% válidos, nenhum erro e nenhum registro com valor deste atributo está vazio e nem nulo.

Consultas

Campanha - Editor do Power Query

Arquivo Página Inicial Transformar Adicionar Coluna Exibição Ferramentas Ajuda

Config. Consulta Layout Visualização dos Dados Consultas Colunas Parâmetros Avançadas Dependências de Consulta Dependências

**NOTADEVOLUCAO**

**QUIDE**

**SOMAQUANT**

**TICKET**

**DRILL**

**REFAT**

**INDICOU**

**NOME**

**DESCRIAO**

100% Válidos 6% Válidos 100% Válidos 100% Válidos 100% Válidos < 1% Válidos 100% Válidos 100% Válidos  
0% Erro  
0% Vazio 94% Vazio 0% Vazio 0% Vazio 0% Vazio 0% Vazio 99% Vazio 0% Vazio 0% Vazio

exclusivo(s) 57 distinto(s), 56 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 1 distinto(s), 0 exclusivo(s) 2 distinto(s), 1 exclusivo(s) 5 distinto(s), 0 exclusivo(s) 9 distinto(s), 0 exclusivo(s)

11 26317 null1 1 S S S S N null LOJA4 GRUPOS  
12 23100 null1 1 S S S S N null LOJA4 GRUPO4  
13 23101 null1 1 S S S S N null LOJA4 GRUPO4  
14 71563 null1 1 S S S S N null LOJA2 GRUPOS  
15 104743 null1 1 S S S S N null LOJA2 GRUPOS  
16 21523 null1 1 S S S S N null LOJA4 GRUPOS  
17 21529 null1 1 S S S S N null LOJA4 GRUPOS  
18 21530 null1 1 S S S S N null LOJA4 GRUPOS  
19 22403 null1 1 S S S S N null LOJA4 GRUPOS

20

Estatísticas da coluna ...

Distribuição de valores

Contagem 1000

Erro 0

Vazio 0

Distintos 9

Exclusivos 0

Cadeia de caracteres vazia 0

Min. GRUPO10

Máx. GRUPO9

GRUPO4 GRUPOS GRUPO9 GRUPO6 GRUPO2 GRUPO8 GRUPO3 GRUPO7

Ativar o Windows  
Acesse Configurações para ativar o Windows.

50 COLUNAS, 199+ LINHAS Criação de perfil de coluna com base nas primeiras 1000 linhas

VISUALIZAÇÃO BAIKADA EM SEGUNDA-FEIRA

POR PTB 26/09/2023

## 6.2. Solução do problema

Depois da carga do database para a ferramenta Power BI, desenvovi um dashboard para atender os objetivos. Demonstro abaixo que o painel foi desenvolvido com conexão no banco de dados no Azure Sql Server (Data Warehouse).

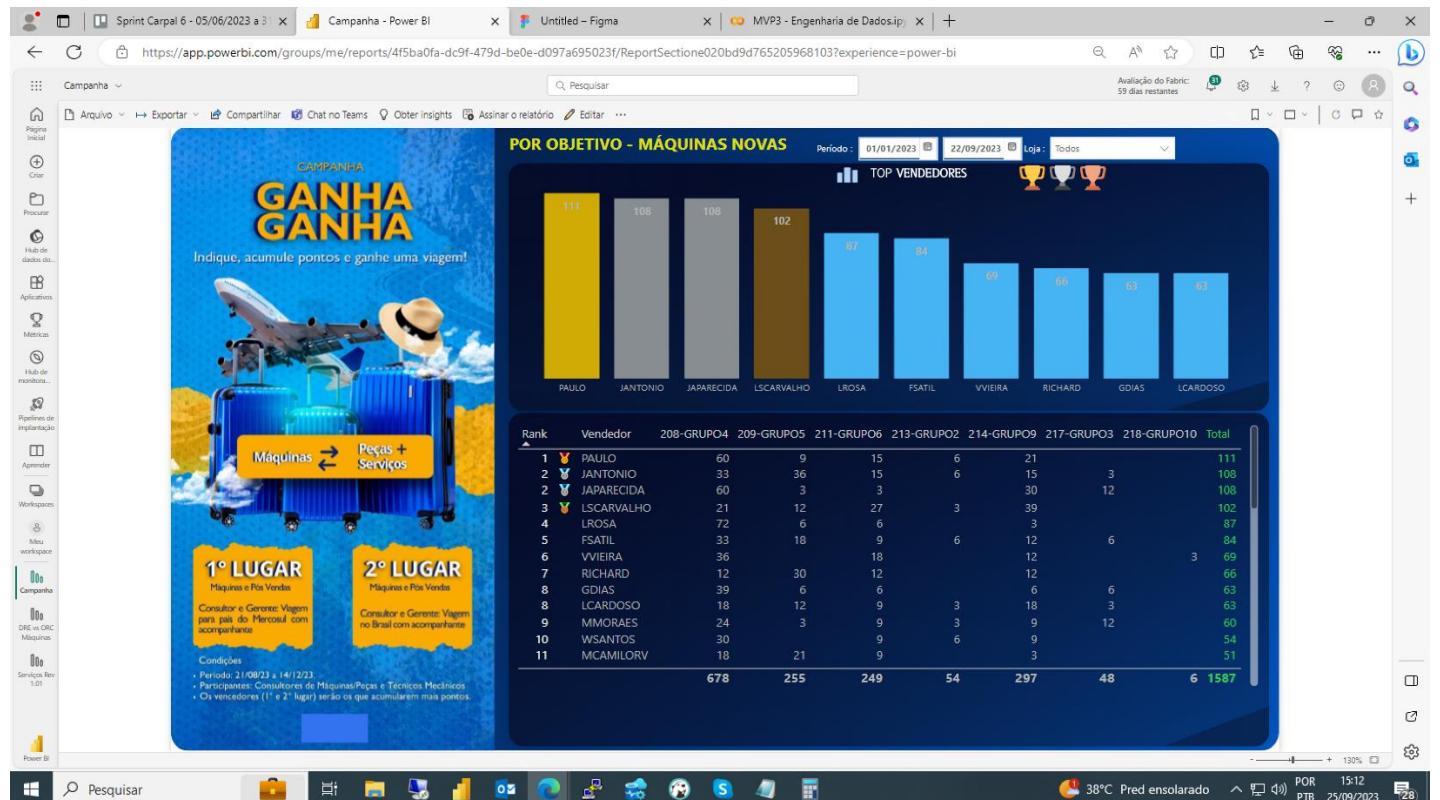
The screenshot shows the Power BI Desktop application window. A modal dialog box titled "Configurações da fonte de dados" (Data Source Settings) is open, displaying a list of data sources. One source, "mvp3sql.database.windows.net:mvp3sql", is selected. The main workspace shows a dashboard titled "POR OBJETIVO - MÁQUINAS NOVAS" with a bar chart and some text. The bottom status bar shows system information like date and time.

Se compararmos os totais das colunas dos grupos de equipamentos vai coincidir com o select(s) número 3 e número 4 feitos no item 4.4 que faz o agrupamento das vendas por grupo. Comparando também o rankeado de número 12 (VENDEDOR: FELIPE V) com o select número 2, veremos que os valores são equivalentes configurando assim a qualidade e acurácia dos dados. Segue abaixo o dashboard criado no POWER BI Desktop.

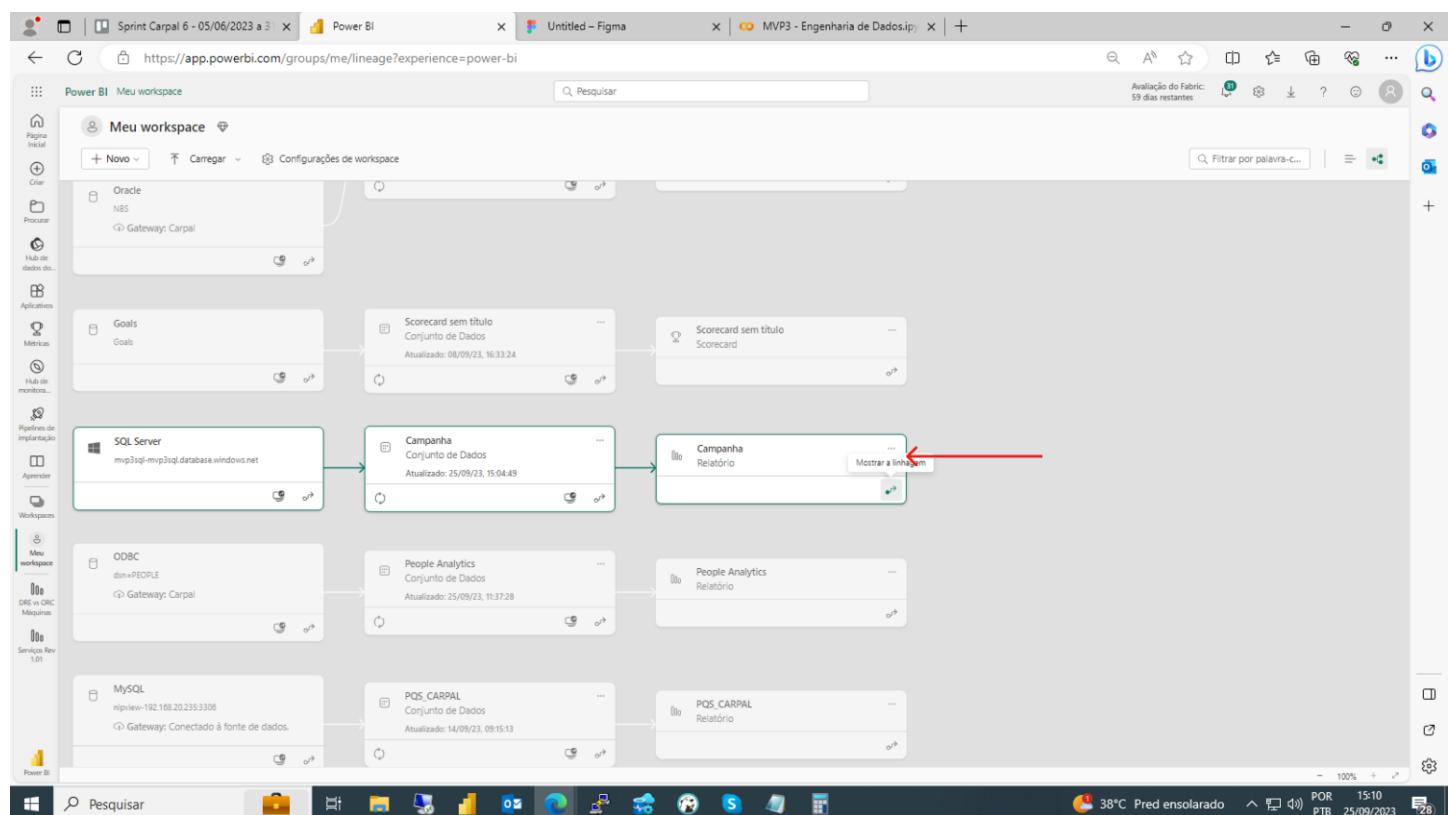
The screenshot shows the Power BI Desktop application window with the completed dashboard. The bar chart and table data now correctly reflect the expected results from the queries. The bottom status bar shows system information like date and time.

Quanto aos objetivos criamos o grid conforme solicitado nos objetivos, satisfazendo aos valores totais por vendedor nas linhas e totais dos grupos de vendas por vendedor nas colunas, também o total geral dos grupos como totalizador das colunas. Os totais gerais batem com os resultados obtidos no Azure Data Studio. O gráfico de barras retrata o total de equipamentos vendidos por vendedor rankeando os mesmos. Chegamos à conclusão que os objetivos foram atendidos.

Dashboard publicado sendo executado.



Na imagem abaixo mostramos a linhagem dos dados pelo POWER BI.



---

## **7. Autoavaliação**

---

Após a realização de todas as etapas do MVP conseguimos atingir todos os itens requeridos no objetivo e retratamos a visualização das informações em um dashboard conforme solicitado. A maior dificuldade foram as configurações dos recursos no Azure e não familiaridade com a ferramenta Purview usada para catalogar os dados.

---