

MVP 3 – Engenharia de Dados

Aluno: Daniel Tinoco Novaes dos Santos

1. Objetivo

Objetivo deste MVP será analisar o comportamento de vendas realizadas em um marketplace de comércio eletrônico brasileiro, no período disponível para análise, a fim de responder as seguintes questões relativas ao negócio:

- 1. Quais os produtos são mais vendidos pelos vendedores no marketplace?*
- 2. Quais são os produtos menos vendidos pelos vendedores no marketplace?*
- 3. Há concentração (sazonalidade) nas vendas em algum período do ano?*
- 4. Há concentração geográfica nas vendas realizadas através do marketplace?*
- 5. Há concentração de vendas em algum vendedor específico?*
- 6. Quais os períodos de maior e menor faturamento no ano?*
- 7. O custo do frete tem impacto na decisão de compra do produto?*

Espera-se ao final desta análise identificar períodos de maior ou menor demanda por um ou mais produtos, fornecendo subsídios para que possa ser possível prever e realizar ações específicas de marketing para promoção de vendas de produtos ou de elevação do volume de vendas em alguma região ou período do ano. Objetivamos também verificar se o custo do frete ou eventuais problemas de atraso na entrega de produtos podem ser fatores determinantes para a decisão de compra pelo cliente, fornecendo subsídios para que a organização possa atuar na correção de problemas logísticos, eliminando deficiências ou diversificando pontos de distribuição.

2. Busca pelos dados

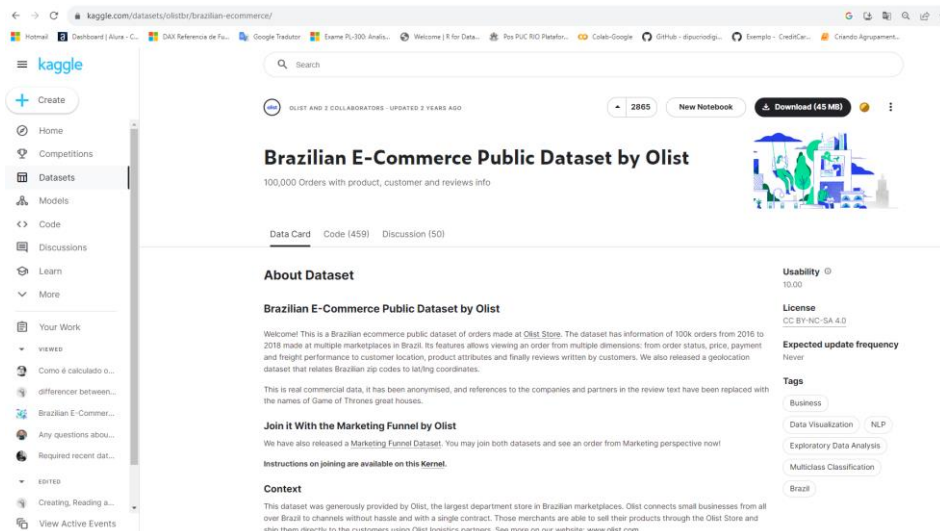
O conjunto de dados utilizado neste MVP foi obtido através de busca no portal Kaggle, uma plataforma dedicada ao aprendizado de ciência de dados mantida pela Google. Trata-se de uma base pública de dados reais de comércio eletrônico disponibilizado por uma empresa de marketplace brasileira, podendo ser acessado através do link: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/>

Este conjunto de dados contém informações de 100 mil pedidos realizados entre os anos de 2016 e 2018. Seus dados permitem visualizar um pedido em várias dimensões: desde o status do pedido, preço, pagamento, o custo do frete até o local de entrega do cliente, atributos do produto e, finalmente, avaliações escritas pelos clientes.

Por se tratar de informações reais, os dados disponibilizados no conjunto foram anonimizados e as referências às empresas e parceiros nas avaliações foram substituídas por nomes utilizados na série de televisão Game of Thrones.

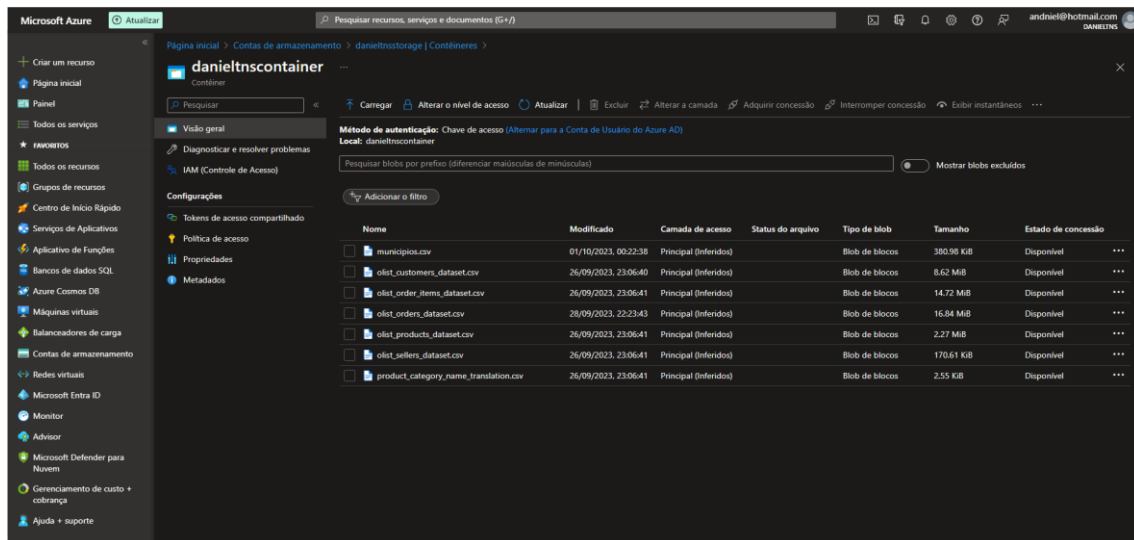
A Olist conecta pequenos negócios de todo o Brasil aos canais de venda sem burocracia e com um único contrato. Os comerciantes podem vender seus produtos por meio da Olist Store e enviá-los diretamente aos clientes usando os parceiros de logística da Olist.

Após a compra de um produto pelo cliente na Olist Store, um vendedor é notificado para atender ao pedido. Assim que o cliente recebe o produto, ou vence a data estimada de entrega, o cliente recebe por e-mail uma pesquisa de satisfação onde pode dar uma nota da experiência de compra e anotar alguns comentários (**Fonte:** Kaggle).



3. Coleta de Dados

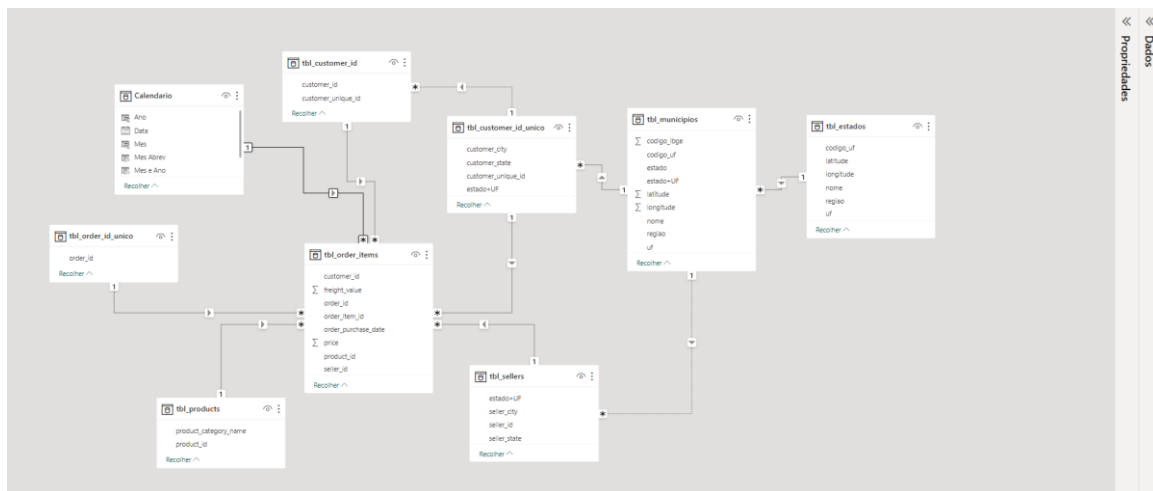
O conjunto de dados utilizado foi coletado no portal Kaggle, baixados para a máquina local e inseridos manualmente em um container criado no storage account do Azure Cloud. Além destes arquivos, foi utilizada uma base de dados contendo as informações de estados e municípios produzidas a partir de dados do IBGE (fonte: <https://github.com/kelvins/Municipios-Brasileiros>), também baixada para a máquina local e inserida no storage account. Complementando o conjunto de dados, foi criada uma tabela diretamente no Banco de Dados SQL do Azure Cloud com um conjunto de dados do IBGE referente aos estados brasileiros, com informações adicionais referentes ao conjunto de dados de municípios.



Container de dados do Azure Cloud contendo os arquivos CSV coletados.

4. Modelagem

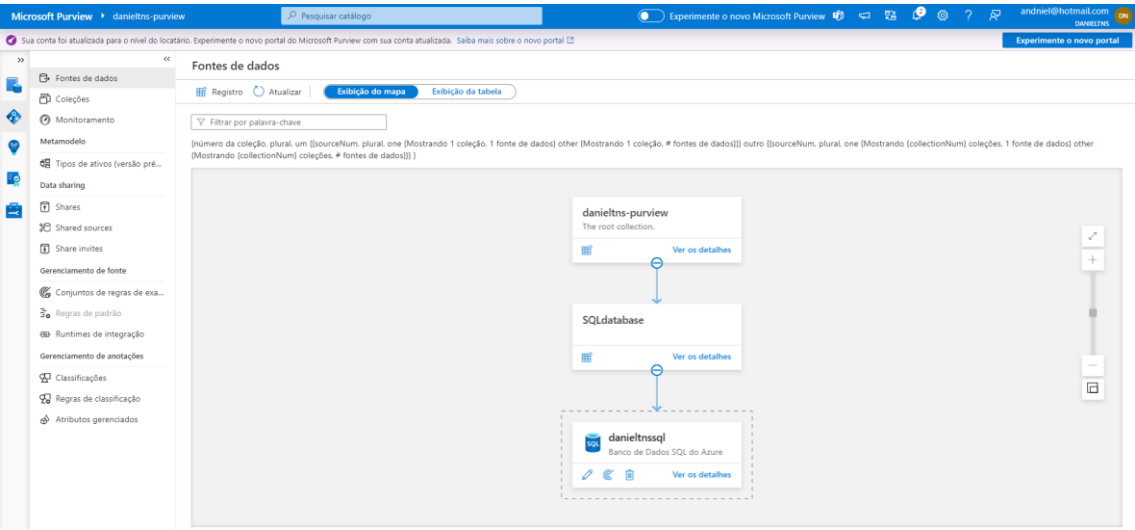
Para realização análise do conjunto de dados utilizado e solução dos problemas de negócios propostos os dados foram modelados em um esquema estrela, com as arquivos armazenados em banco de dados relacional.



Esquema Estrela construído no Power BI

Para construções do Catálogo de Dado foi utilizado o Azure Purview, ferramenta disponibilizada pelo Azure Cloud para Governança da dados. Nas imagens abaixo é

disponibilizada a hierarquia de dados obtidas no Purview a partir do Banco de Dados SQL, bem como a descrição de cada tabela e seus respectivos domínios.



Fonte de Dados – Microsoft Purview

Coleções	
Coleções são grupos de itens, como fontes de dados e ativos, que são organizados juntos no Mapa de Dados. Saiba mais sobre coleções.	
Filtrar por palavra-chave	
danieltns-purview	
SQLdatabase	
Visão Geral Atribuições de função	
Atualizado em 30 de setembro de 2023, 00:54 por Daniel Novaes	
Descrição	
Nenhuma descrição para esta coleção	
Caminho da coleção	
danieltns-purview	
SQLdatabase	
Criado em	
30 de setembro de 2023, 00:54 por Daniel Novaes	

Resultado do escaneamento de dados realizado pelo Microsoft Purview

- Tabela customer_id

Microsoft Purview

danielnrs-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andniet@hotmail.com

09/10/2023

Sua conta foi atualizada para o nível do locatário. Experimente o novo portal do Microsoft Purview com sua conta atualizada. Saiba mais sobre o novo portal

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danielnrsq[>

tbl_customer_id

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆ @

↑ ↓

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:08 por Daniel Noveas

Descrição do ativo

Este conjunto de dados contém informações sobre o cliente. Utilizado para identificação do identificador do cliente (customer_id) gerados nos pedidos e para identificação único do cliente (customer_unique_id).

No conjunto de dados utilizado, para cada pedido é atribuído a um customer_id exclusivo. Isso significa que o mesmo cliente receberá IDs diferentes para pedidos diferentes. O objetivo de ter um customer_unique_id no conjunto de dados é permitir identificar os clientes únicos no conjunto de dados.

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema

Não há classificações para este ativo.

Nome totalmente qualificado

mssql://danielnrsq.database.windows.net/danielnrsq/dbo/tbl_customer_id

Caminho da coleção

danielnrs-purview

SQLdatabase

Hierarquia

danielnrsq.database.windows.net

SQL Server do Azure

danielnrsq

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_customer_id

tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danielnrsq[>

tbl_customer_id

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆ @

↑ ↓

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir Editar colunas

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:08 por Daniel Noveas

Filtrar por nome

Mostrando 2 de 2 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
customer_id			customer_id	nvarchar	ID do cliente para cada pedido realizado.
customer_unique_id			customer_unique_id	nvarchar	ID único do cliente no banco de dados.

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:08 por Daniel Noveas

Pesquisar ativos ou processos

tbl_customer_id_unico

Tabela do SQL do Azure

Servidor: danielnrsq.database.windows.net

Banco de d...: danielnrsq

Esquema: dbo

Mudar para o ativo

tbl_customer_id

Tabela do SQL do Azure

Servidor: danielnrsq.database.windows.net

Banco de d...: danielnrsq

Esquema: dbo

Linhagem dos dados – tbl_customer_id

- Tabela customer_id_unico

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltrssql >

tbl_customer_id_unico

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:44 por Daniel Novaes

Descrição do ativo

Este conjunto de dados contém informações sobre o cliente com seu identificador único, a cidade de residência e o estado de residência.

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema (2)

Country/Region

World Cities

Nome totalmente qualificado

mssql://danieltrssql.database.windows.net/danieltrssql/dbo/tbl_customer_id_unico

Caminho da coleção

...

danieltrssql-view

SQLdatabase

Hierarquia

danieltrssql.database.windows.net

SQL Server do Azure

danieltrssql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_customer_id_unico

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltrssql >

tbl_customer_id_unico

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:44 por Daniel Novaes

Filtrar por nome

Mostrando 3 de 3 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
customer_unique_id			customer_unique_id	nvarchar	ID único do cliente no banco de dados.
customer_city	World Cities		customer_city	nvarchar	Cidade brasileira onde reside o cliente.
customer_state	Country/Region		customer_state	nvarchar	Estado brasileiro de residência do cliente.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltrssql >

tbl_customer_id_unico

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:44 por Daniel Novaes

Pesquisar ativos ou processos

Tabela do SQL do Azure

tbl_customer_id_unico

Servidor: danieltrssql.database.windows.net

Banco de... danieltrssql

Esquema: dbo

tbl_customer_id

Linhagem de dados – tbl_customer_id_unico

- Tabela estados

Microsoft Purview

danieltns-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andtnet@hotmail.com

DANIELTNS

Sua conta foi atualizada para o nível do locatário. Experimente o novo portal do Microsoft Purview com sua conta atualizada. Saiba mais sobre o novo portal

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnsql >

tbl_estados

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:46 por Daniel Noveas

Descrição do ativo

Este conjunto de dados consolida informações sobre os estados brasileiros (fonte de dados IBGE).

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema (1)

Geolocation (Lat/Lon)

Nome totalmente qualificado

mssql:/danieltnsql.database.windows.net/danieltnsql/dbo/tbl_estados

Caminho da coleção

danieltns-purview

SQLdatabase

Hierarquia

danieltnsql.database.windows.net

SQL Server do Azure

danieltnsql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_estados

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnsql >

tbl_estados

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir Editar colunas

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:46 por Daniel Noveas

Filtrar por nome

Mostrando 6 de 6 itens

- Tabela municípios

Microsoft Purview

danieltns-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andtnet@hotmail.com

DANIELTNS

Sua conta foi atualizada para o nível do locatário. Experimente o novo portal do Microsoft Purview com sua conta atualizada. Saiba mais sobre o novo portal

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnsql >

tbl_municipios

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar Selecionar para edição em massa Solicitar acesso Atualizar Excluir

Abrir no Power BI Desktop

Visão Geral Propriedades Esquema Linhagem Contatos Relacionado

Atualizado em 1 de outubro de 2023 às 15:46 por Daniel Noveas

Descrição do ativo

Este conjunto de dados consolida informações sobre os municípios brasileiros (fonte de dados IBGE).

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema (1)

Geolocation (Lat/Lon)

Nome totalmente qualificado

mssql:/danieltnsql.database.windows.net/danieltnsql/dbo/tbl_municipios

Caminho da coleção

danieltns-purview

SQLdatabase

Hierarquia

danieltnsql.database.windows.net

SQL Server do Azure

danieltnsql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_municipios

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnssql >

tbl_municipios

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆🔍⬆⬇

✎ Editar

👤 Selecionar para edição em massa

📄 Solicitar acesso

🔄 Atualizar

🗑 Excluir

⚙ Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:46 por Daniel Novaes

🔍 Filtrar por nome

Mostrando 5 de 5 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
codigo_ibge			codigo_ibge	int	Código em números atribuídos pelo IBGE aos municípios do Brasil.
nome				nvarchar	Nome do município.
latitude	📍 Geolocation (Lat/Lon)			real	latitude do município (posição geográfica).
longitude	📍 Geolocation (Lat/Lon)			real	latitude do município (posição geográfica).
codigo_uf			codigo_uf	int	Código em números da unidade da federação atribuído ao Estado Brasileiro pelo L...

- Tabela order_id_unico

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnssql >

tbl_order_id_unico

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆🔍⬆⬇

✎ Editar

👤 Selecionar para edição em massa

📄 Solicitar acesso

🔄 Atualizar

🗑 Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 16:31 por Daniel Novaes

Descrição do ativo

Tabela com identificador único de compra sem registros duplicados.

Atributos gerenciados

🔍 Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema

Não há classificações para este ativo.

Nome totalmente qualificado

mssql://danieltnssql.database.windows.net/danieltnssql/dbo/tbl_order_id_unico

Caminho da coleção

...

danieltns-purview

SQLdatabase

Hierarquia

danieltnssql.database.windows.net

SQL Server do Azure

danieltnssql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_order_id_unico

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnssql >

tbl_order_id_unico

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆🔍⬆⬇

✎ Editar

👤 Selecionar para edição em massa

📄 Solicitar acesso

🔄 Atualizar

🗑 Excluir

⚙ Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

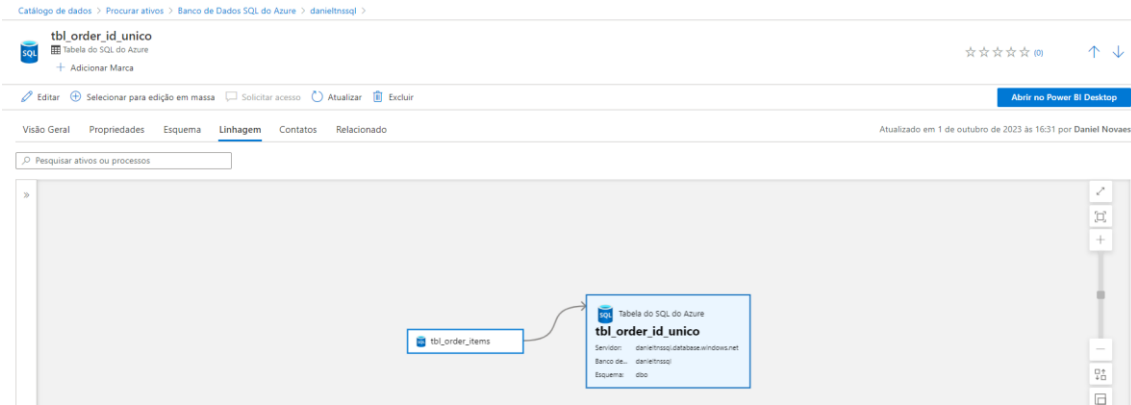
Relacionado

Atualizado em 1 de outubro de 2023 às 16:31 por Daniel Novaes

🔍 Filtrar por nome

Mostrando 1 de 1 item

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
order_id			order_id	nvarchar	Identificador único do pedido.



Linhagem de Dados – tbl_order_id_unico

- Tabela order_items

Microsoft Purview

danieltns-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andnias@hotmail.com

DANIELTNS

Sua conta foi atualizada para o nível do locatário. Experimente o novo portal do Microsoft Purview com sua conta atualizada. Saiba mais sobre o novo portal

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnssql >

tbl_order_items

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:49 por Daniel Noveas

Descrição do ativo

Este conjunto de dados consolida informações referentes às compras realizadas pelos clientes.

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema

Não há classificações para este ativo.

Nome totalmente qualificado

mssql://danieltnssql.database.windows.net/danieltnssql/dbo/tbl_order_items

Caminho da coleção

danieltns-purview

SQL.database

Hierarquia

danieltnssql.database.windows.net

SQL Server do Azure

danieltnssql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbl_order_items

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danieltnssql >

tbl_order_items

Tabela do SQL do Azure

+ Adicionar Marca

☆☆☆☆☆

↑ ↓

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

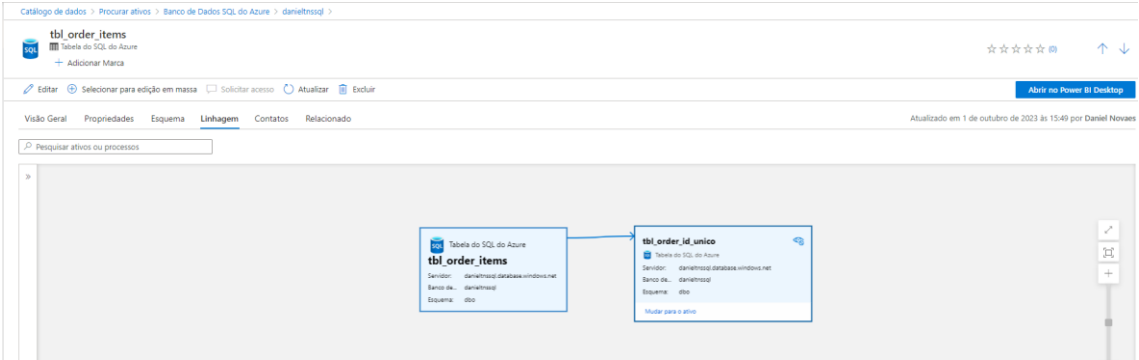
Relacionado

Atualizado em 1 de outubro de 2023 às 15:49 por Daniel Noveas

Filtrar por nome

Mostrando 8 de 8 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
order_id			order_id	nvarchar	Identificador único do pedido.
order_item_id			order_item_id	nvarchar	Número sequencial que identifica o número de itens incluídos em um mesmo pedido. Tipo de dado texto.
product_id			product_id	nvarchar	Identificador exclusivo do produto. Tipo de dado texto.
seller_id			seller_id	nvarchar	Identificador único do vendedor. Tipo de dado texto.
price				real	Preço do produto. Tipo de dados float (2 casas decimais). Valor máximo 6.735. Valor mínimo 0.85.
freight_value				real	Valor do frete do produtos (se um pedido tiver mais de um item o valor do frete é dividido entre os itens). Tipo de dados float (com 2 casas decimais). Valor máximo 22,30 e valor mínimo.
customer_id			customer_id	nvarchar	ID do cliente para cada pedido realizado. Tipo de dado texto.
order_purchase...			order_purchase_date	date	Data e hora da compra realizada pelo cliente. Tipo de dado data.



Linhagem de Dados – tbl_order_items

- Tabela products

Microsoft Purview

danilms-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andrei@thermal.com

danilms

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danilmssql >

tbi_products

Tabela do SQL do Azure

+ Adicionar Marca

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:50 por Daniel Novaes

Descrição do ativo

Este conjunto de dados consolida informações sobre os produtos vendidos.

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema

Não há classificações para este ativo.

Nome totalmente qualificado

mssql://danilmssql.database.windows.net/danilmssql/dbo/tbi_products

Caminho da coleção

danilms-purview

SQLdatabase

Hierarquia

danilmssql.database.windows.net

SQL Server do Azure

danilmssql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbi_products

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danilmssql >

tbi_products

Tabela do SQL do Azure

+ Adicionar Marca

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:50 por Daniel Novaes

Filtrar por nome

Mostrando 2 de 2 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
product_id			product_id	nvarchar	Identificador exclusivo do produto.
product_category_name			product_category_name	nvarchar	Nome da categoria do produto.

- Tabela sellers

Microsoft Purview

danilms-purview

Pesquisar catálogo

Experimente o novo Microsoft Purview

andrei@thermal.com

danilms

Experimente o novo portal

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danilmssql >

tbi_sellers

Tabela do SQL do Azure

+ Adicionar Marca

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:51 por Daniel Novaes

Descrição do ativo

Este conjunto de dados consolida informações sobre os vendedores, os municípios e estados onde se localizam.

Atributos gerenciados

Filtrar por nome de atributo

Mostrar os atributos sem um valor

Nenhum Atributo para este ativo.

Classificações

Não há classificações para este ativo.

Classificações de esquema (2)

Country/Region

World Cities

Nome totalmente qualificado

mssql://danilmssql.database.windows.net/danilmssql/dbo/tbi_sellers

Caminho da coleção

danilms-purview

SQLdatabase

Hierarquia

danilmssql.database.windows.net

SQL Server do Azure

danilmssql

Banco de Dados SQL do Azure

dbo

Esquema do Azure SQL

tbi_sellers

Tabela do SQL do Azure

Termos do glossário

Não há termos do glossário para este ativo.

Catálogo de dados > Procurar ativos > Banco de Dados SQL do Azure > danilmssql >

tbi_sellers

Tabela do SQL do Azure

+ Adicionar Marca

Editar

Selecionar para edição em massa

Solicitar acesso

Atualizar

Excluir

Editar colunas

Abrir no Power BI Desktop

Visão Geral

Propriedades

Esquema

Linhagem

Contatos

Relacionado

Atualizado em 1 de outubro de 2023 às 15:51 por Daniel Novaes

Filtrar por nome

Mostrando 3 de 3 itens

Nome da coluna	Classificações	Rótulo de conf...	Termos do glossário	Tipo de dados	Descrição da coluna
seller_id			seller_id	nvarchar	Identificador único do vendedor.
seller_city	World Cities		seller_city	nvarchar	Nome da cidade de localização do vendedor.
seller_state	Country/Region		seller_state	nvarchar	Estado de localização do vendedor.

Não foi possível a realização de linhagem automatizada de dados referente as tabelas *tbl_estado*, *tbl_municipios*, *tbl_order_items*, *tbl_customer_id*, *tbl_products* e *tbl_sellers* diretamente do Banco de Dados SQL por limitação do aplicativo MS Purview, que não reconheceu a origem dos dados após a ingestão no banco de relacional através do Data Factory. Todos dados apresentados tem como origem primária os arquivos CSV ingeridos no Servidor SQL e Pipeline de dados construídos com os arquivos no MS Data Factory.

5. Carga

A extração, transformação e carga dos dados foi realizada utilizando o Azure Data Factory e Power Query, este último usando a versão nativa embarcada no aplicativo Power BI na preparação dos dados para análise.

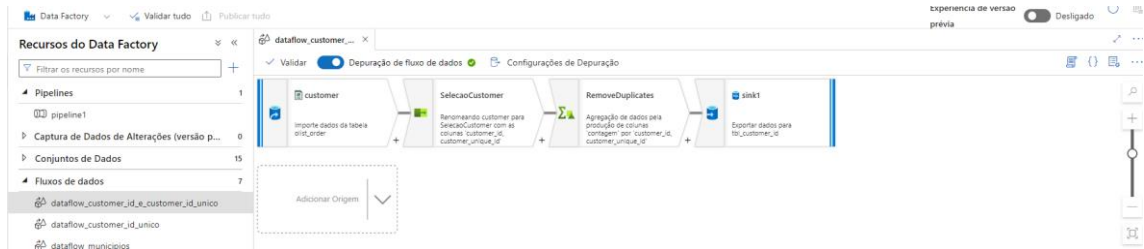
A ingestão de dados no Banco de Dados SQL do Azure foi realizada através da ferramenta Data Factory. Foram construídos 6 diferentes fluxos de dados para ingestão das tabelas no servidor. Neste fluxo de dados foram excluídas colunas que não seriam necessárias para análise, a definição dos tipos de dados e a remoção de linhas duplicadas nas tabelas que compuseram o esquema construído.

Fluxo de dados - *tbl_customer_id*

Para construção da *tbl_customer_id* foram realizadas as seguintes etapas no fluxo de dados:

- Importação da tabela *olist_customer_dataset.csv*
- Seleção das colunas *customer_id* e *customer_unique_id*
- Remoção das tuplas duplicadas utilizando um processo de agregação

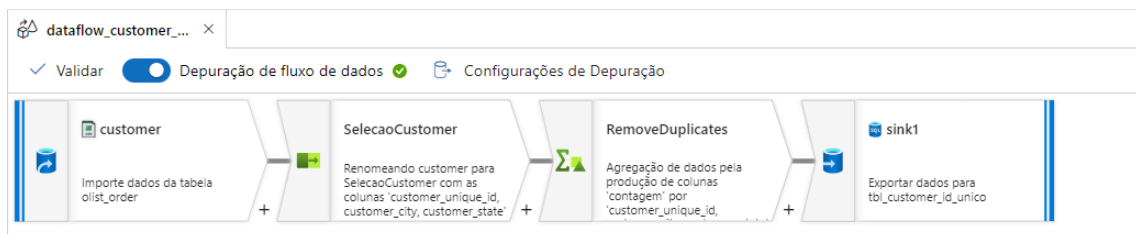
- Ingestão no banco de dados



Fluxo de dados - tbl_customer_id_unico

Para construção da *tbl_customer_id_unico* foram realizadas as seguintes etapas no fluxo de dados:

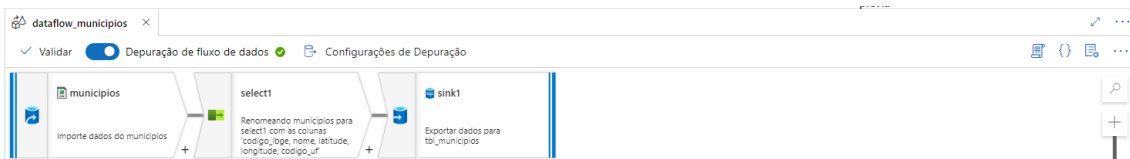
- Importação dos dados da tabela *olist_order_dataset.csv*
- Seleção das colunas *customer_unique_id*, *customer_city* e *customer_state*
- Remoção das tuplas duplicadas utilizando um processo de agregação
- Ingestão no banco de dados



Fluxo de dados - tbl_municipios

Para construção da *tbl_municipios* foram realizadas as seguintes etapas no fluxo de dados:

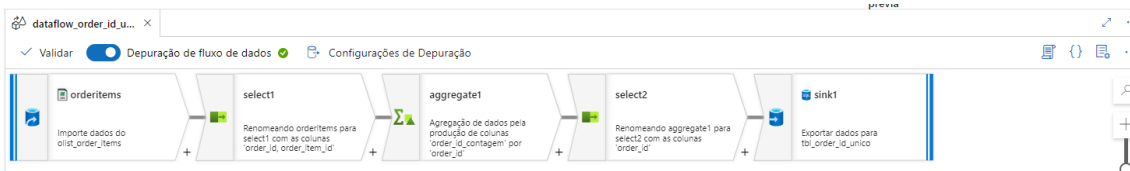
- Importação dos dados da tabela *municipios.csv*
- Seleção das colunas *codigo_ibge*, *nome*, *latitude*, *longitude* e *código_uf*
- Ingestão no banco de dados



Fluxo de dados - *tbl_order_id_unico*

Para construção da *tbl_order_id_unico* foram realizadas as seguintes etapas no fluxo de dados:

- Importação da tabela *olist_order_items_dataset.csv*
- Seleção das colunas *order_id* e *order_item_id*
- Remoção das tuplas duplicadas utilizando um processo de agregação
- Seleção da coluna *order_id*
- Ingestão no banco de dados

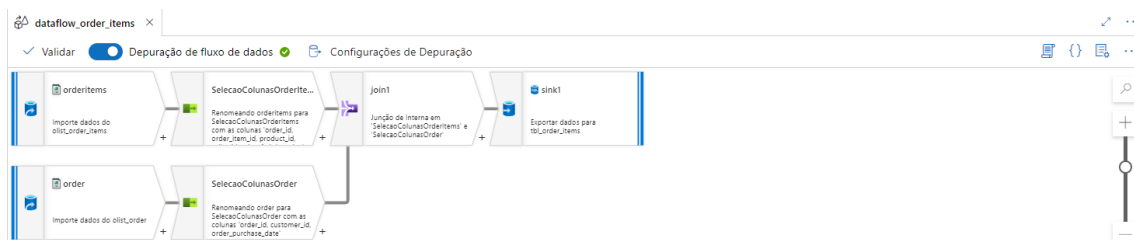


Fluxo de dados - *tbl_order_items*

Para construção da *tbl_order_items* foram realizadas as seguintes etapas no fluxo de dados:

- Importação da tabela *olist_order_items_dataset.csv*
- Seleção das colunas *order_id*, *order_item_id*, *product_id*, *seller_id*, *price* e *freight_value*
- Importação da tabela *olist_order_dataset.csv*

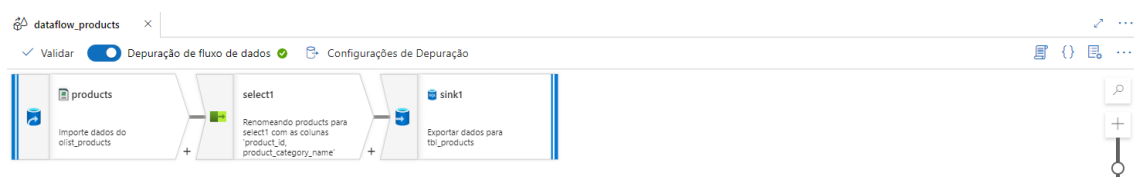
- Seleção das colunas *order_id*, *customer_id*, *order_purchase_timestamp*
- Renomeada a coluna *order_purchase_timestamp* para *order_purchase_date*
- Realizado uma Junção (*inner*) com condição de junção sendo as colunas *order_id* de ambas as tabelas iguais
- Ingestão no banco de dados



Fluxo de dados - *tbl_products*

Para construção da *tbl_products* foram realizadas as seguintes etapas no fluxo de dados:

- Importação da tabela *olist_products_dataset.csv*
- Seleção das colunas *product_id* e *product_category_name*
- Ingestão no banco de dados



Fluxo de dados - *tbl_sellers*

Para construção da *tbl_sellers* foram realizadas as seguintes etapas no fluxo de dados:

- Importação da tabela *olist_sellers_dataset.csv*
- Seleção das colunas *seller_id*, *seller_city* e *seller_state*
- Ingestão no banco de dados



Após a construção e validação dos Fluxos de Dados foi realizada a ingestão de dados propriamente dita no Banco de Dados SQL através da montagem e execução do Pipeline de Dados (abaixo).

Nome da atividade	Status da atividade	Tipo de atividade	Início da execução	Duração	Runtime de integração	Propriedades do usuário	ID de execução de atividade	Log
dataflow_sellers	Bem-sucedido	Fluxo de dados	10/1/2023, 4:24:20 PM	2m 51s	AutoResolveIntegrationRl		1025a5cd-a23f-430e-b269-93d49a74d55b	
dataflow_products	Bem-sucedido	Fluxo de dados	10/1/2023, 4:24:20 PM	4m 20s	AutoResolveIntegrationRl		06bb208c-2376-422d-9e04-a230cd6ae5d	
dataflow_order_items	Bem-sucedido	Fluxo de dados	10/1/2023, 4:24:20 PM	3m 53s	AutoResolveIntegrationRl		00c05790-c334-4326-b0c0-4ddb8d24f8ae	
dataflow_municipios	Bem-sucedido	Fluxo de dados	10/1/2023, 4:24:20 PM	3m 56s	AutoResolveIntegrationRl		bc0a9d31-754b-4be1-b318-1353bb5c989f	
dataflow_customer_id_unico	Bem-sucedido	Fluxo de dados	10/1/2023, 4:24:20 PM	3m 24s	AutoResolveIntegrationRl		14ef9f6-3c6f-4077-861a-6bab0061c3dd	

Recursos do Data Factory

Filtrar os recursos por nome	
Pipelines	1
pipeline1	
Captura de Dados de Alterações (versão p...	0
Conjuntos de Dados	15
Fluxos de dados	7
dataflow_customer_id_e_customer_id_unico	
dataflow_customer_id_unico	
dataflow_municipios	
dataflow_order_id_unico	
dataflow_order_items	
dataflow_products	
dataflow_sellers	
Power Query	0

Fluxo de Dados no Azure Data Factory

Após a importação dos dados para o aplicativo Power BI, foi realizada a construção do esquema estrela mostrado anteriormente, com a definição dos relacionamentos entre as tabelas descritos abaixo, além da construção de uma tabela dimensão de calendário, a fim de auxiliar nas análises a serem realizadas.

Tabela de Relacionamentos

Nome Tabela	Chave PK	Referência externa	Chave FK
tbl_customer_id	customer_id	tbl_order_items	customer_id
Calendario	Data	tbl_order_items	order_purchase_date
tbl_order_id_unico	order_id	tbl_order_items	order_id
tbl_products	product_id	tbl_order_items	product_id
tbl_sellers	seller_id	tbl_order_items	seller_id
tbl_customer_id_unico	customer_unique_id	tbl_order_items	customer_id
tbl_municipios	estado+UF	tbl_customer_id_unico	estado+UF
tbl_estados	codigo_uf	tbl_municipios	codigo_uf

6. Análise

▪ Qualidade dos dados

O conjunto de dados não apresentou grandes problemas em relação aos domínios dos dados disponibilizados. Entre atributos de valor numérico *price*, *freight_value* e *order_item_id* não verificamos valores extremos que demonstrasse não serem valores possíveis para os atributos avaliados. Nestes atributos não observamos valores vazios ou nulos que precisassem de tratamento específico. Em relação ao atributo *order_purchase_date* (data de compra do cliente) também não observamos datas que estivessem em formato inadequado ou muito distantes do intervalo disponível no conjunto de dados. Por tratar-se de um conjunto de dados disponibilizados no Portal

Kaggle, com as primeiras versões do dataset publicadas há cinco anos, este conjunto já pode ter tido muitos dos seus ruídos e inconsistências tratados ao longo do tempo.

Em relação aos atributos *customer_unique_id* observamos problemas para inclusão da chave primária de relação entre as tabelas utilizadas em razão de haver código identificador de clientes duplicados, porém com atributos de cidade e estado diferentes. Como a proposta deste atributo é disponibilizar um identificador único do cliente, que pudesse ser utilizado identificar o cliente no banco de dados da empresa, haver tuplas diferentes com o atributo *customer_unique_id* repetido foi um problema que precisou ser tratado do ponto de vista semântico, para não haver repetidos. Como não foi disponibilizada data de origem destas informações no *dataset* original e como não havia como saber qual das tuplas era a mais recente ou a correta, optou por escolher aleatoriamente uma única tupla para que o *customer_unique_id* fosse único e pudesse ser utilizado como chave primaria, tratamento realizado no Pipeline de Dados executado no Data Factory.

Os atributos de identificação *customer_id*, *order_id*, *product_id* e *seller_id* foram anonimizados e por isso não foi possível avaliar se qualidade dos dados para estes atributos, como a existência de um ID válido, por exemplo, estava comprometida. Nas demais colunas, relacionadas à categoria dos produtos (*product_category_name*) e estados e municípios (*seller_city*, *seller-state*, *customer_city*, *customer_state*) também não observamos ocorrências que comprometessem a qualidade dos dados utilizados de uma forma geral. As tabelas com os dados disponibilizados pelo IBGE, com informações de estados e municípios brasileiros, estavam com seus atributos também sem ocorrências relacionadas à qualidade dos dados. Apenas foi necessário o tratamento para supressão de assentos e caracteres especiais nos nomes das cidades e estados, a fim de que fosse

possível realizar a relação entre estas tabelas e informações de cidade e estado das tabelas *tbl_seller* e *tbl_customer_id_único*.

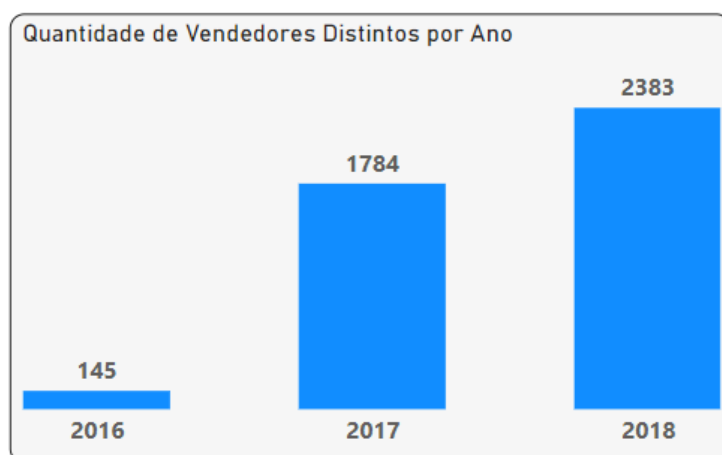
- **Solução do Problema**

- 1. Quais os produtos são mais vendidos pelos vendedores no marketplace?*

O conjunto de dados utilizados possui registros históricos de vendas de produtos no Marketplace da Olist entre os dias 04/09/2016 e 03/09/2028, compreendendo um período total de 24 meses de dados. Quando analisamos as vendas totais nos anos de 2016, 2017 e 2018 percebemos que a ausência de dados referentes aos 8 primeiros meses do ano de 2016 e os últimos 4 meses do ano de 2018 refletem nos resultados quando buscamos identificar os produtos mais vendidos ano a ano. Conforme dados abaixo, é possível notar que, diferente dos anos de 2017 e 2018, onde os produtos mais vendidos foram os produtos de cama, mesa e banho, em 2016 os produtos da categoria de móveis e decoração foram os mais comercializados. Esta diferença pode ser resultado de um número mais baixo de vendedores durante o primeiro ano de dados disponíveis. Nesse período, a plataforma contava apenas com 145 vendedores distintos comercializando produtos em seu marketplace. Nos anos de 2017 e 2018 a quantidade de vendedores comercializando produtos pelo portal aumentou significativamente, passando a totalizar 1.784 e 2.383 vendedores distintos, respectivamente.

Top 5 categorias mais vendidas por ano

Categoria do Produto	2016	Categoria do Produto	2017	Categoria do Produto	2018
moveis_decoracao	104	cama_mesa_banho	6.302	cama_mesa_banho	7.347
beleza_saude	59	moveis_decoracao	5.776	beleza_saude	6.778
perfumaria	37	esporte_lazer	4.735	informatica_acessorios	5.930
brinquedos	29	beleza_saude	4.244	moveis_decoracao	5.660
market_place	29	informatica_acessorios	3.917	utilidades_domesticas	5.315



2. Quais são os produtos menos vendidos pelos vendedores no marketplace?

Entre os produtos com pior desempenho de vendas no período avaliado estão alimentos, calçados, roupas femininas e masculinas, fraldas e livros em 2016 e produtos da categoria pc gamer para os anos de 2017 e 2018.

Top 5 categorias menos vendidas por ano

Categoria do Produto	2016	Categoria do Produto	2017	Categoria do Produto	2018
alimentos	1	pc_gamer	4	pc_gamer	7
fashion_calçados	1	portateis_casa_forno_e_cafe	4	fashion_esporte	6
fashion_roupa_feminina	1	artesanato	3	la_cuisine	5
fashion_roupa_masculina	1	fraldas_higiene	3	fashion_roupa_infanto_juvenil	3
fraldas_higiene	1	seguros_e_servicos	2	cds_dvds_musicais	1
livros_interesse_geral	1				
livros_tecnicos	1				

Para análise das questões 1 e 2 foram excluídos os itens em que não foram localizados o nome da sua categoria através do seu identificador, após a relação das tabelas, por não podermos inferir a que categorias estes produtos pertencem.

3. Há concentração (sazonalidade) nas vendas em algum período do ano?

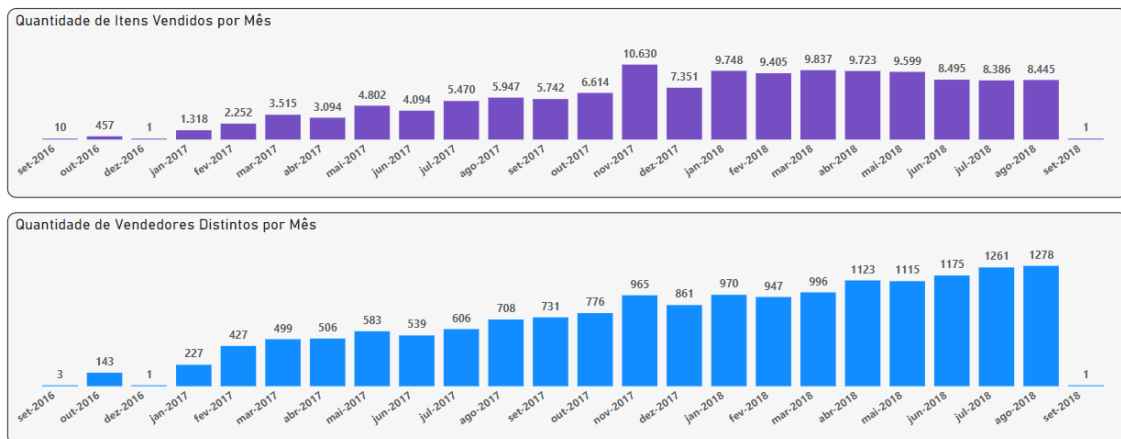
Analisando a totalidade dos produtos vendidos ao longo dos meses, não foi possível detectar efeitos de sazonalidades nas vendas realizadas no marketplace da Olist. Podemos perceber que ao longo dos meses de 2017 e 2018 houve na verdade um aumento constante na quantidade de itens vendidos pelo portal, com pico de produtos vendidos no mês de novembro de 2017, movimento que pode ser explicado por ser um mês de grandes ações de marketing, como a Blackfriday.

Quantidade de itens vendidos por mês e ano



Podemos observar nos gráficos abaixo que a quantidade de itens vendidos ficou relativamente estável a partir do mês de janeiro de 2018, mesmo com a continuidade crescente de vendedores comercializando através do portal, indicando que a média de produtos vendidos por vendedor começou a apresentar declínio a partir deste mesmo mês.

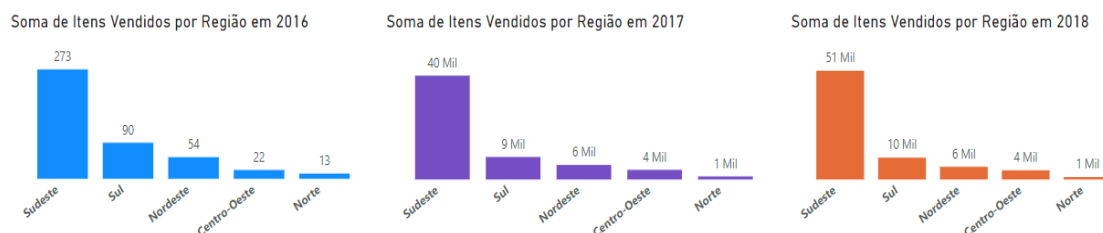
Quantidade total de itens vendidos e de vendedores distintos por mês e ano

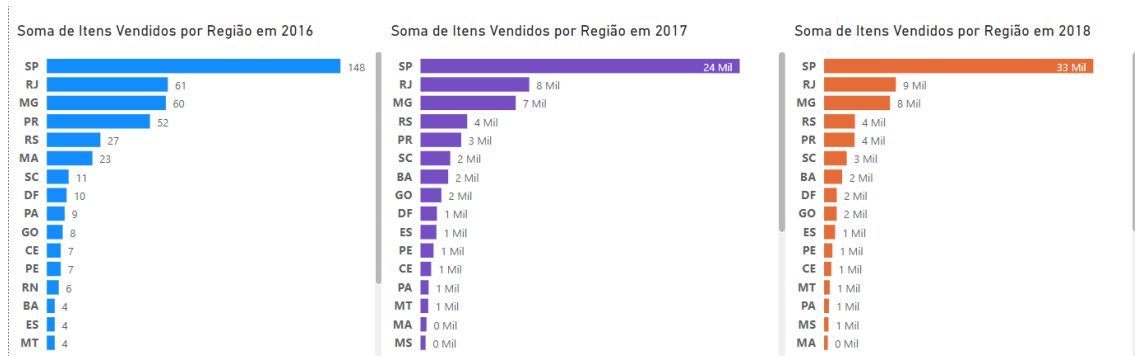


4. Há concentração geográfica nas vendas realizadas através do marketplace?

Quando avaliamos a quantidade de item vendidos por Região do país observamos uma concentração expressiva das vendas realizadas na Região Sudeste em todo período avaliado. Os estados de São Paulo, Rio de Janeiro e Minas Gerais foram os estados com maior quantidade de produtos vendidos em todos os anos disponíveis no conjunto de dados.

Quantidade itens vendidos por Região e Estado





5. Há concentração de vendas em algum vendedor específico?

Não verificamos grandes concentrações de vendas por vendedor no conjunto de dados analisado. Há uma grande quantidade de vendedores comercializando produtos nos anos de 2017 e 2018 (1.784 e 2.383 vendedores, respectivamente), com os maiores vendedores responsáveis por 2,61% do total de itens vendidos em 2017 e 2,15% do total de itens vendidos em 2018. Para o ano de 2016 este percentual se mostra um pouco mais elevado, com 6,62% da quantidade total de itens vendidos por apenas um vendedor. Porém, neste primeiro ano a quantidade de vendedores cadastrados para comercialização de seus produtos através do portal Olist era de apenas 145, ou seja, apenas 10% dos vendedores que comercializaram produtos no ano de 2017, o que contribui para a concentração do volume de vendas naquele ano.

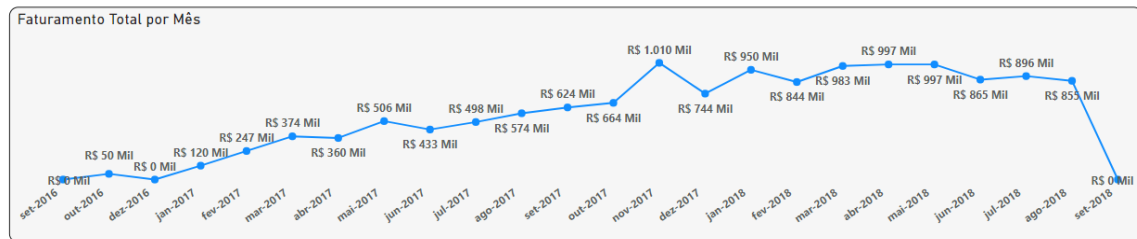
Concentração de itens vendidos por vendedor

% de Participação no Total de Itens Vendidos em 2016		% de Participação no Total de Itens Vendidos em 2017		% de Participação no Total de Itens Vendidos em 2018	
seller_id	% Total de Itens Vendidos	seller_id	% Total de Itens Vendidos	seller_id	% Total de Itens Vendidos
620c87c171fb2a6dd6e8bb4dec959fc6	6,62%	1f50f920176fa81dab994f9023523100	2,61%	1025f0e2d44d7041d6cf58b6550e0bfa	2,15%
391fc631aebcf3004804e51b40bcf1e	5,77%	4a3ca9315b744ce9f8e9374361493884	2,33%	955fee9216a65b617aa5c0531780ce60	2,11%
ecccf2b993b34a3bf033cc5d1dcdc69	5,13%	cc419e0650a3c5ba77189a1882b7556a	2,17%	1f50f920176fa81dab994f9023523100	1,78%
024b564ae893ce8e9bfa02c10a401ece	4,49%	6560211a19b47992c3666cc4a7e94c0	1,76%	7c67e1448b00f6e969d365cea6b010ab	1,62%
cca3071e3e9bb7d12640c9f9be2301306	4,49%	7c67e1448b00f6e969d365cea6b010ab	1,61%	6560211a19b47992c3666cc4a7e94c0	1,62%
ed859002ad59dbf8cf3602696a6c3000	3,85%	cca3071e3e9bb7d12640c9f9be2301306	1,48%	da8622b14eb17ae2831f4ac5b9dab84a	1,48%
897060da8b9a21f655304d50fd935913	3,42%	7a67c85e85bb2ce8582c35f2203ad736	1,33%	8b321bb669392f5163d04c59e235e066	1,27%
46dc3b2cc0980fb8ec44634e21d2718e	2,56%	da8622b14eb17ae2831f4ac5b9dab84a	1,32%	ea8482cd71df3c1969d7b9473ff13abc	1,22%
d101c6da914ae3d53f7bee42283f2fe2	2,35%	3d871de0142ce09b7081e2b9d1733cb1	1,21%	4a3ca9315b744ce9f8e9374361493884	1,15%
0c8380b62e38e8a1e6adbeba7eb9688c	2,14%	1025f0e2d44d7041d6cf58b6550e0bfa	1,17%	4869f7a5dfa277a7dca6462dcf3b52b2	1,02%
5b179e9e8cc7ab6fd113a46ca584da81	1,92%	d2374cbcb3ca4ab1086534108cc3ab7	0,82%	7d13fca15225358621be4086e1eb0964	0,82%
82bd0703a4aefdb599e5b5fdaed378fb	1,92%	f8db351d8c4c4c22c6835c19a46f01b0	0,82%	e9779976487b77c6d4ac45f75ec7afe9	0,80%
0336182e1b3e92f029d5354832045fdf	1,71%	46dc3b2cc0980fb8ec44634e21d2718e	0,81%	1835b56ce799e6a4dc4eddc053f04066	0,76%
Total	100,00%	Total	100,00%	Total	100,00%

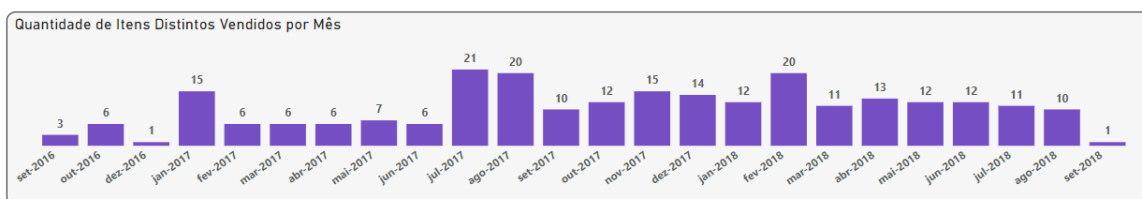
6. Quais os períodos de maior e menor faturamento no ano?

Verificamos que o pico de maior faturamento realizado pelo Portal Olist foi no mês de novembro de 2017, onde houve um volume total de vendas de R\$ 1,010 Milhão. Podemos observar no gráfico abaixo um volume crescente vendas ou longo dos meses, com estabilidade entre os meses de janeiro de 2018 e agosto de 2018, meses que apresentaram um valor médio mensal de vendas de R\$ 912 Mil. Em relação ao mês de menor valor de faturamento, verificamos que este mês foi dezembro de 2016, com valor de vendas total abaixo de mil reais. O mês em questão é o terceiro mês da série histórica disponível, podendo indicar que no início das atividades do marketplace a empresa ainda apresentava poucos vendedores e falta de regularidade nas comercialização de produtos. Nesse período a quantidade de itens diferentes vendidos era muito baixa, pouco diversificada, podendo ser uma das razões para o baixo faturamento.

Faturamento total por mês e ano



Quantidade de itens distintos por mês e ano

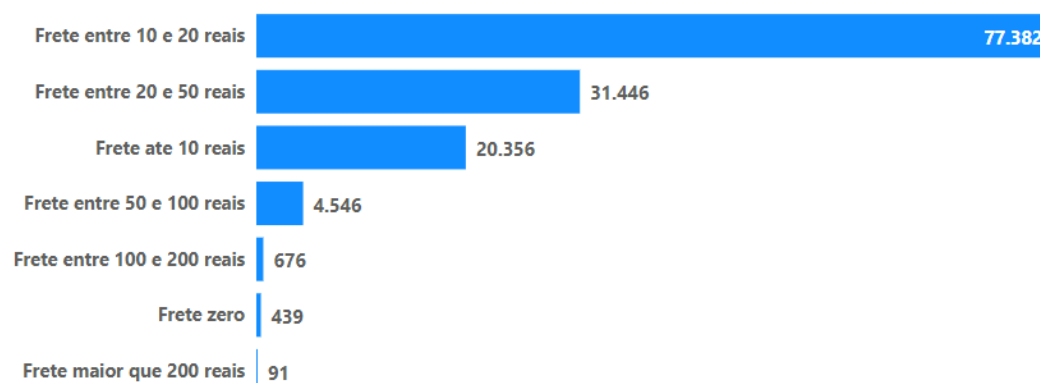


7. O custo do frete tem impacto na decisão de compra do produto?

Em relação ao custo do frete, vemos que os itens mais vendidos no marketplace da Olist apresentaram um valor baixo de frete, com valores de frete cobrados para a entrega dos produtos entre 10 e 20 reais. Conforme os valores de frete aumentam a quantidade de itens vendidos também diminui, indicando que o valor do frete para a entrega do produto pode sim ser um fator de compra decisivo para o consumidor. Uma vez que verificamos há uma concentração de vendas de itens na região sudeste, podemos inferir que um custo menor de frete pode estar relacionado a logística mais barata de entrega naquela região e que a diferenciação de locais para distribuição nas áreas onde o volume de vendas é menor pode contribuir para potencialização dos negócios em outras áreas.

Quantidade de itens Vendidos por Faixa de Frete

Quantidade de Itens Vendidos por Faixa de Frete



7. Conclusão

Ao longo das análises realizadas verificamos que o marketplace Olist apresentou um bom crescimento no valor total de vendas e na quantidade de itens vendidos do início da série histórica observada até o último trimestre de 2017, quando houve uma estabilidade. Há uma concentração e dependência elevada de vendas realizadas na região Sudeste do país, o que pode indicar uma dificuldade da empresa em viabilizar vendas em outras regiões. Esta dificuldade pode estar relacionada a um custo maior de frete para entrega dos produtos, o que observamos na quantidade baixa de itens vendidos com fretes elevados, ou ainda em uma capacidade menor de penetração junto ao consumidor de outras localidades, fato que não foi objeto do presente estudo.

8. Autoavaliação

O estudo foi realizado com o objetivo de, utilizando-se uma base real de vendas de um marketplace brasileiro, realizar a extração, transformação e cargas de dados em um ambiente de cloud, com análise das questões de negócios levantadas no início do estudo, após o procedimento de ETL e modelagem dos dados.

Optamos por utilizar neste estudo o ambiente de cloud da Microsoft Azure, fazendo uso das ferramentas Azure Data Factory, Azure Data Studio, Azure Purview e Power BI, este último por ser uma ferramenta de Business Intelligence foi utilizado para análise dos dados. Apesar da integração entre as ferramentas da Microsoft, não conseguimos realizar a integração entre o Data Factory e Purview de forma que, quando da carga e ingestão de dados, fosse possível obter o Catálogo de Dados e Linhagem do Dados de forma automatizada, mesmo que as origens dos datasets utilizados fossem o servidor SQL do Azure Cloud e as autorizações de função e implementações de segurança relativas ao aplicativo fossem efetuadas. Os testes de conexão foram realizados com êxito, mas sem que o a linhagem dos dados e o catálogo de dados fossem disponibilizados. Os avisos informados no MS Purview indicam que esta funcionalidade ainda está em “preview”, o que pode sugerir uma melhora no futuro. O Data Studio foi pouco utilizado nesse relatório, usado apenas para verificação através consultas nas tabelas disponíveis no servidor após a ingestão de dados, o que pode ser mais bem explorado em estudos futuros. Além das dificuldades técnicas, o preparo do ambiente de nuvem para a realização do ETL consumiu muito tempo, entre configuração do ambiente e aprendizagem da ferramenta, o que reduziu o tempo disponível para análise das questões de negócio.

Em relação os objetivos de análise dos dados, outras perguntas de negócios podem ser realizadas com o conjunto de dados usado, como a avaliação das reviews dos clientes

sobre as compras e sobre produtos que estavam disponibilizadas no dataset original. Optou-se nesse momento por não as utilizar em função do tempo disponível. As perguntas de negócios propostas foram respondidas, mas em todos os questionamentos levantados é possível aprofundar ainda mais a análise para fornecer um diagnóstico mais completo do negócio, investigar ainda mais as razões e sinalizações de consumo realizadas pelo cliente e fornecer os *insights* que sirvam de subsídio para a ampliação de negócios e para o planejamento estratégico da organização.