

Loja Sulamérica

Projeto da

Disciplina de BI

Prof. Anderson Nascimento
prof.anderson@ica.ele.puc-rio.br

Componentes do Projeto:

André Luis Mendes Teixeira
Gabriela de Camargo Santa Rosa
Mariana Fernandes Coy

Histórico de Versões

Data	Versão	Descrição	Autor	Aprovado por
01/10/2021	1.0	Este é um documento de elaboração do projeto de BI para a Loja Sulamérica expondo os modelos transacionais, multidimensionais e relatórios, modificação na capa e nos caps. 1, 3 e 9	André Teixeira	Gabriela C. Santa Rosa Mariana F. Coy
28/10/2021	2.0	Inclusão de informações do modelo multidimensional	André Teixeira	Gabriela C. Santa Rosa Mariana F. Coy
03/11/2021	3.0	Errata do modelo multidimensional e consulta de geração de tabela fato	André Teixeira	Gabriela C. Santa Rosa Mariana F. Coy
05/11/2021	4.0	Errata no texto da introdução e descrição de estudo de casos	André Teixeira	Gabriela C. Santa Rosa Mariana F. Coy
06/11/2021	5.0	Revisão geral do relatório e conclusão	Gabriela C. Santa Rosa	André Teixeira
08/11/2021	6.0	Erratas no documento e inclusão dos painéis do Dashboard	Mariana F. Coy	Gabriela C. Santa Rosa Mariana F. Coy

Sumário

1 Introdução.....	3
2 Estudo de Caso	4
2.1 Descrição do Estudo de Caso.....	4
3 Descrição do Modelo Transacional	4
3.1 Fonte 1 – Diagrama de modelo de dados físico.....	4
3.2 Fonte 2 – Entrevista.....	5
3.3 Descrição do projeto ETL Transacional	6
4 Modelo Multidimensional	8
4.1 Descrição	8
5 Projeto de ETL.....	9
5.1 Descrição do Projeto de ETL	9
6 Dashboard	12
6.2 Telas do Dashboard.....	12
7 Conclusão.....	15
8 Arquivos.....	15
8.1 pentaho_transformations.zip	15
8.2 scripts.zip	15
8.3 Modelos_power_architect_projeto.zip	15
8.4. Archive.zip	15
8.5 PBI-Projetov2.PBIX.....	15

1 Introdução

Este documento tem por finalidade coletar, analisar e definir as principais necessidades do projeto do estudo de caso Lojas Sulamérica. O documento procura analisar dados de ordens de pedidos de produtos entre os períodos de setembro de 2016 a outubro de 2018, demonstrando os principais KPIs (Faturamento, Lucro, Markup e Custo Total) e criando uma visão genérica para análise dos resultados da empresa e também uma visão específica para as áreas Financeiro e Customer Success, identificando pontos de melhoria para a empresa, provendo informações que suportem o patrocinador para que ações sejam tomadas para o gerenciamento do negócio.

2 Estudo de Caso

2.1 Descrição do Estudo de Caso

A Loja Sulamérica é uma empresa do ramo varejista criada em 2016 no Rio de Janeiro onde fica a sua sede e que atua em todo o território nacional, através de seu market place disponível na web e aplicativo.

Com o aumento exponencial das vendas, os diretores sentiram a necessidade de obter informações importantes sobre os segmentos das lojas e verificar onde a empresa está mais atuante e onde possa ter um maior crescimento em suas vendas.

Por isso a Loja Sulamérica solicitou apoio da consultoria de nossa equipe de BI para elaboração de relatórios e painéis gerenciais que auxilie a empresa a entender melhor o seu negócio analisando as vendas registradas nos períodos do quarto trimestre de 2016 até o terceiro trimestre de 2018.

O projeto tratará as ordens de pedido dos produtos adquiridos como fato, levando em consideração as dimensões de itens de produtos vendidos, clientes por estados e as formas de pagamento.

Por ordem da Lei Geral de Proteção de Dados (LGPD), os nomes de clientes, nomes de vendedores foram preservados em sua base transacional, sendo identificados apenas por Ids criptografados. Sendo assim, teremos uma análise de granularidade maior, não obtendo informações pessoais.

De acordo com os dirigentes da empresa contratante dos serviços de BI, são esperados relatórios do tipo: maiores Estados consumidores, maiores pontos de vendas, tipos de produtos mais vendidos, formas de pagamento mais utilizadas, sempre que possível, mostrando o período temporal.

A base de origem são itens, produtos, clientes, vendedores, ordens de pedido, avaliações de clientes em formato Excel e espera-se, com o projeto, a construção de um Data Warehouse, além de dashboards com os dados mais relevantes mencionados pelos dirigentes da Companhia para melhor avaliação de todo o processo de venda, onde a empresa poderá facilmente ter as informações que apoiem em suas tomadas de decisões.

3 Descrição do Modelo Transacional

3.1 Fonte 1 – Diagrama de modelo de dados físico.

Como parte da análise e entendimento do funcionamento do sistema transacional, foi coletado o diagrama do modelo de dados físico do sistema de ordem de pedidos da Companhia.

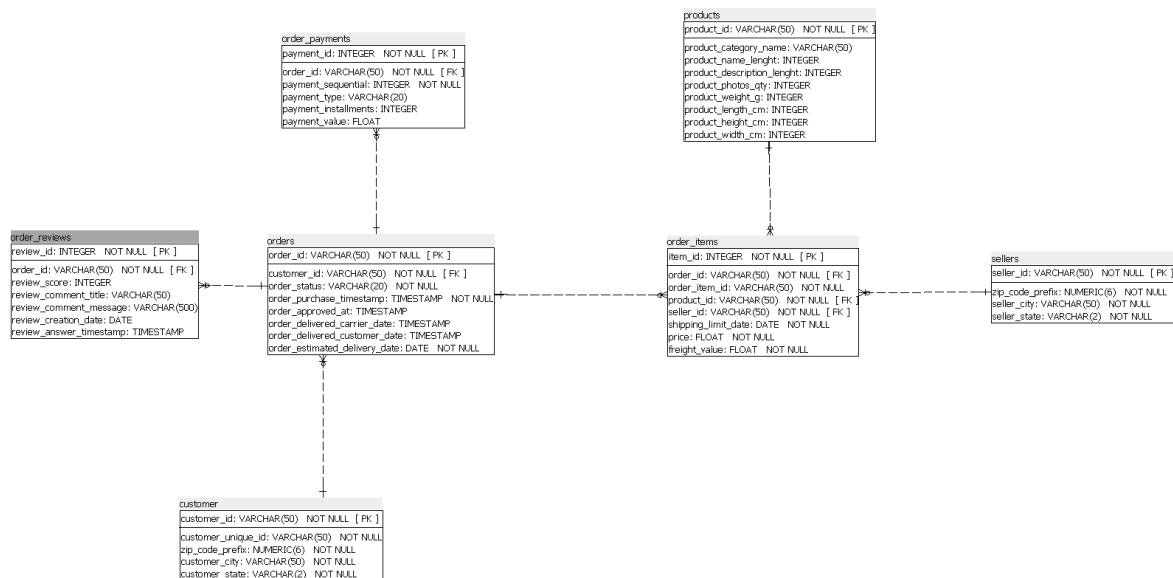


Figura 1 – Diagrama do modelo físico do Sistema de ordem de pedidos da Loja Sulamérica

3.2 Fonte 2 – Entrevista.

Foi realizada uma entrevista com os stakeholders da empresa para o entendimento das necessidades do negócio. Os pontos fundamentais elicitados durante as entrevistas foram fundamentais para o entendimento da necessidade de análise do negócio.

A organização possuiu todas as fontes de dados em arquivos em formato excel (.xlsx) com as seguintes descrições:

olist_customers_dataset.xlsx

Descrição: lista de clientes que fizeram compras na loja.

olist_order_items_dataset

Descrição: lista de itens vendidos

olist_order_payments_dataset

Descrição: lista de ordem de pagamentos dos pedidos

olist_order_reviews_dataset

Descrição: lista com as avaliações feitas após os pedidos realizados

olist_orders_dataset

Descrição: lista dos pedidos

olist_products_dataset

Descrição: lista de produtos das lojas.

olist_sellers_dataset

Descrição: lista de vendedores

3.3 Descrição do projeto ETL Transacional

Foi realizada a extração dos dados de acordo com o modelo transacional representado na Figura 1 utilizando a ferramenta Pentaho Data Integration (PDI), onde o destino da base está na base Projeto do Postgres.

Com as sete listas apresentadas, foram criados sete processos de carga realizadas na seguinte ordem para obedecer às regras de integridade do modelo físico proposto:

- 1) Clientes (carga_transacional_customers.ktr)

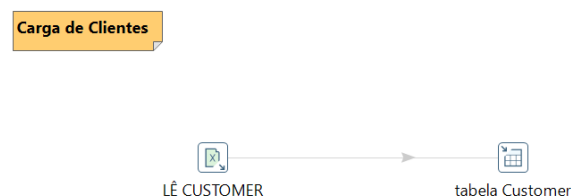


Figura 2 – Carga transacional customers

- 2) Vendedores (carga_transacional_sellers.ktr)



Figura 3 – Carga transacional sellers

3) Produtos (carga_transacional_products.ktr)



Figura 4 – Carga transacional products

4) Pedidos (carga_transacional_orders.ktr)



Figura 5 – Carga transacional orders

5) Itens de pedidos (carga_transacional_orderitens.ktr)

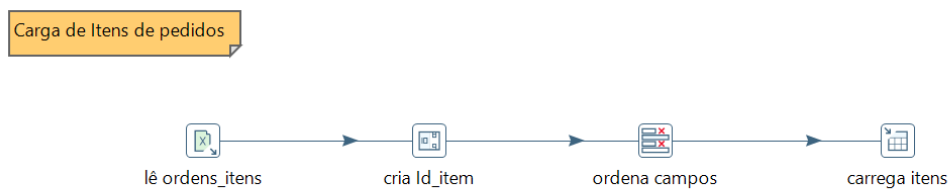


Figura 5 – Carga transacional orders_itens

6) Ordens de pagamentos (carga_transacional_orderpayments.ktr)

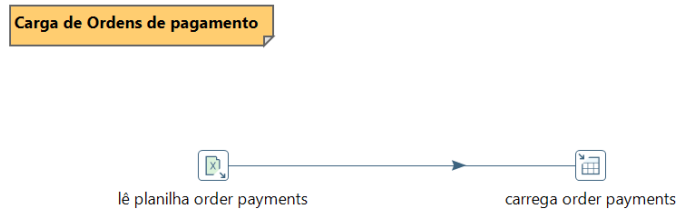


Figura 6 – Carga transacional order_payments

7) Avaliações (carga_transacional_orderreviews.ktr)

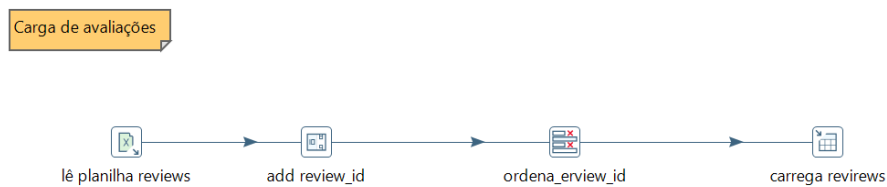


Figura 7 – Carga transacional order_reviews

4 Modelo Multidimensional

4.1 Descrição

Esta seção apresenta o modelo estrela (star schema) do estudo de caso da Loja Sulamérica.

Este modelo foi desenvolvido de acordo com o resultado do processo de análise das necessidades do cliente, e o resultado está representado na Figura 2.

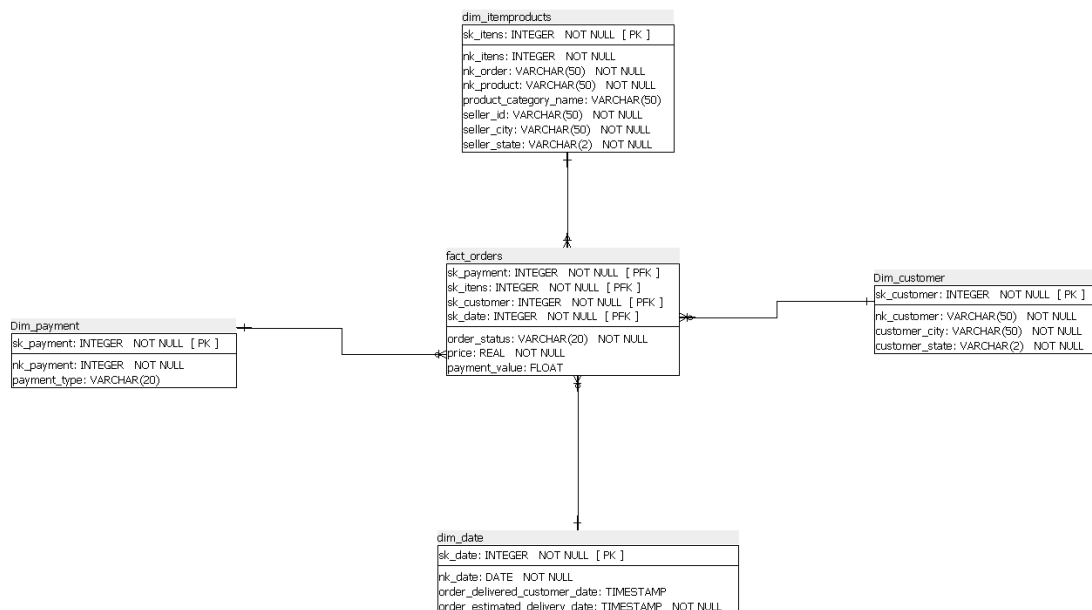


Figura 8 – Modelo Multidimensional

5 Projeto de ETL

5.1 Descrição do Projeto de ETL

O projeto de ETL foi realizado extraindo dados da base transacional realizada e demonstrada no item 3 deste documento.

Foi realizada a extração dos dados de acordo com o modelo multidimensional representado na Figura 2. No processo foi utilizado o software Pentaho Data Integration (PDI).

Para a dimensão tempo, foi extraída dos campos de data de ordem de compra (order_purchase_timestamp). O processo de carga no PDI chamado de carga_multidim_date.ktr, apresentado no diagrama abaixo:



Figura 9 – ETL dim_date

Para a carga da dimensão cliente, foi realizado a extração de cidades e estados de onde foram os compradores mantendo o sigilo de informações pessoais de clientes. O processo e carga do PDI chamado de carga_multidim_customers.ktr, apresentado no diagrama abaixo:



Figura 10 – ETL dim_customers

A transformação string operations foi utilizada nesta carga para tratar a string do campo customer_city, deixando a cadeia de strings em minúsculo e posteriormente as letras iniciais em maiúsculo. Em sequência, foi utilizado a transformação replace in string para tratar as preposições do campo customer_city, deixando-as em minúsculo (ex: a operação anterior que transformou rio de janeiro em **Rio De Janeiro**, passa agora a ser transformado para Rio **de** Janeiro).

Para a carga da dimensão itens / produtos, buscou-se o relacionamento entre os itens vendidos, as categorias de produtos e as cidades e estados dos vendedores. O processo de carga do PDI chamado de carga_multidim_itemproducts.ktr, apresentado no diagrama abaixo:

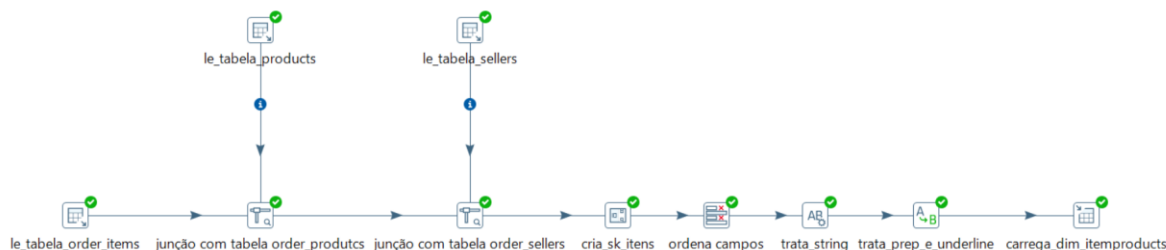


Figura 11 – ETL dim_itemproducts

A transformação string operations foi utilizada nesta carga para tratar as strings dos campos seller_city e product_category_name, deixando a cadeia de strings em minúsculo e posteriormente as letras iniciais em maiúsculo. Em sequência, foi utilizada a transformação replace in string para tratar as preposições do campo seller_city para deixá-las em letra minúscula (da mesma forma que foi feito no ETL da dim_customer), e tratar o underline do campo product_category_name, onde na análise dos dados foi verificado que os dados deste campo estavam utilizando underline para separar a cadeia de caracteres, no lugar de espaço.

Para a carga da dimensão pagamentos, filtrou-se por dados de forma de pagamentos e valores. O processo de carga do PDI chamado de `carga_multidim_payments.ktr`, apresentado no diagrama abaixo:



Figura 12 – ETL dim_payments

A transformação `replace in string` foi utilizada para tratar o campo `payment_type`. Na análise dos dados foi verificado que os dados deste campo estavam utilizando underline para separar a cadeia de caracteres, no lugar de espaço. Portanto esta transformação foi utilizada para substituir o underline por espaço (ex: `credit_card` se transformou em `credit card`). Em sequência, foi utilizado a transformação `string operations` para tratar a string do mesmo campo, deixando a cadeia de strings em minúsculo e posteriormente as letras iniciais em maiúsculo.

Para a carga de dados da tabela fato foi usado o processo de carga do PDI chamado `carga_multidim_fact_v5.ktr`, apresentado no diagrama abaixo:

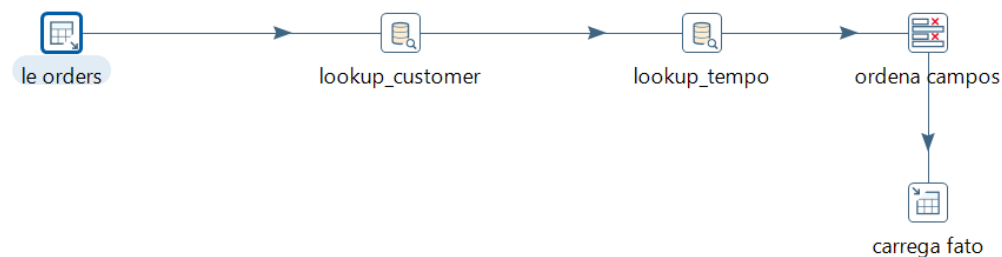


Figura 13 – ETL dim_itemsproducts

Observação importante: Para a carga da tabela `fact_orders`, temos que considerar a exclusão de ordens de pedido em que apresentam status de indisponível (“unavailable”) e cancelado (“canceled”), uma vez que não se obtém itens de produtos com esses status.

Portanto, no fluxo `carga_multidim_fact_v5.ktr` no item “`le orders`” temos a seguinte restrição:

```
select o.order_id, o.order_status, i.item_id, p.payment_id,o.customer_id,
o.order_purchase_timestamp,i.price,p.payment_value
from orders o,
order_items i,
order_payments p
where o.order_id=i.order_id
and i.order_id= p.order_id
and o.order_status not in ('unavailable','canceled')
```

Alguns dados não foram carregados para o Data Warehouse por não serem necessários para análises futuras da área responsável.

Todas as transformações feitas no PDI estão no arquivo pentaho_transformations.zip. Todos os Scripts utilizados estão documentados no arquivo scripts.zip.

6 Dashboard

6.1 Descrição da Elaboração

O Dashboard foi construído com a ferramenta Power BI, da Microsoft. Esta ferramenta foi escolhida devido à fácil utilização por ter uma interface amigável entre usuário e aplicação, além da possibilidade de tê-la disponível gratuitamente.

6.2 Telas do Dashboard

O Dashboard criado possui 3 abas para atender às solicitações das análises da Loja Sulamérica.

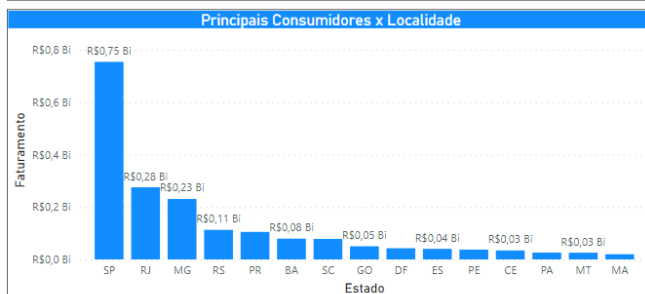
Na aba Visão Geral, destaca-se o Faturamento e apresentam-se as vendas por categoria de produtos em quantidade e em valor. Além disso, as localidades com maior faturamento são evidenciadas tanto no mapa quanto no gráfico de barras. Todos os dados são filtráveis por ano, conforme solicitado pelo cliente.

Na aba Financeiro destacam-se os KPIs Lucro e Custo Total. Também são apresentadas as informações de Quantidade Vendida, Faturamento, Custo Total e Markup (margem do produto, calculada através do Faturamento/Custo Total) por categoria de produto. Essas visões são complementadas com um gráfico do faturamento por método de pagamento e análise de valor de vendedores por estado.

Por fim, na aba Customer Success destacamos os pedidos que não foram finalizados em quantidade e em faturamento, além de identificar as localidades com maiores casos problemáticos.

Dashboard - Sulamérica

Categorias mais vendidas		
Categoria dos Produtos	Quantidade Vendida	Total Vendas (R\$)
Cama mesa banho	11115	R\$171.125.808,00
Beleza saude	9670	R\$165.373.045,00
Esporte lazer	8641	R\$138.136.323,00
Moveis decoracao	8334	R\$142.478.252,00
Informatica acessorios	7827	R\$157.154.381,00
Utilidades domesticas	6964	R\$108.656.532,00
Relogios presentes	5991	R\$142.171.528,00
Telefonia	4545	R\$48.505.615,00
Total	112650	R\$2.015.617.558,00



Ano		
2016	2017	2018

R\$2,02 Bi
Faturamento



Figura 14 – Dashboard Loja Sulamérica – Visão Geral

Dashboard - Sulamérica

Categorias mais vendidas				
Categoria dos Produtos	Quantidade Vendida	Total Vendas (R\$)	Custo Total (R\$)	Markup
Telefonia fixa	264	R\$20.520.350,00	R\$5.854.248,00	3,51
Sinalizacao e seguranc	199	R\$6.848.805,00	R\$2.158.481,00	3,17
Bebidas	379	R\$7.136.045,00	R\$2.280.457,00	3,13
Fraldas higiene	39	R\$418.123,00	R\$153.879,00	2,72
Moveis escritorio	1691	R\$64.636.686,00	R\$28.577.896,00	2,26
Casa conforto 2	30	R\$171.054,00	R\$77.317,00	2,21
Artigos de natal	153	R\$1.897.384,00	R\$881.385,00	2,15
Flores	33	R\$221.301,00	R\$111.004,00	1,99
Moveis decoracao	8334	R\$142.478.252,00	R\$76.270.474,00	1,87
Total	112650	R\$2.015.617.558,00	R\$1.410.576.700,00	1,43

R\$605,04 Mi
Lucro

R\$1,41 Bi
Custo

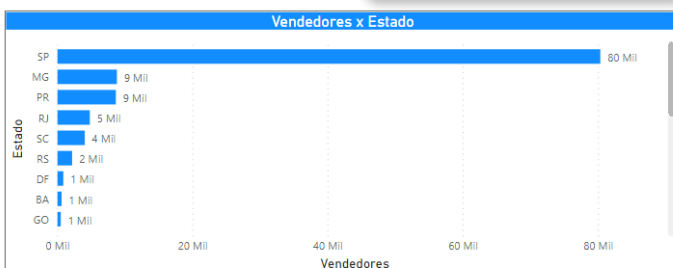
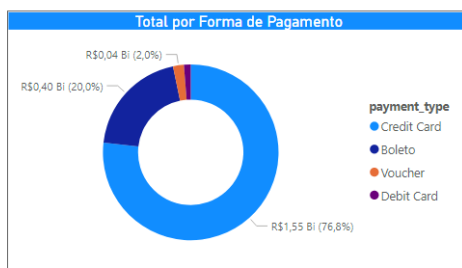


Figura 15 – Dashboard Loja Sulamérica – Visão Financeiro



Dashboard - Sulamérica

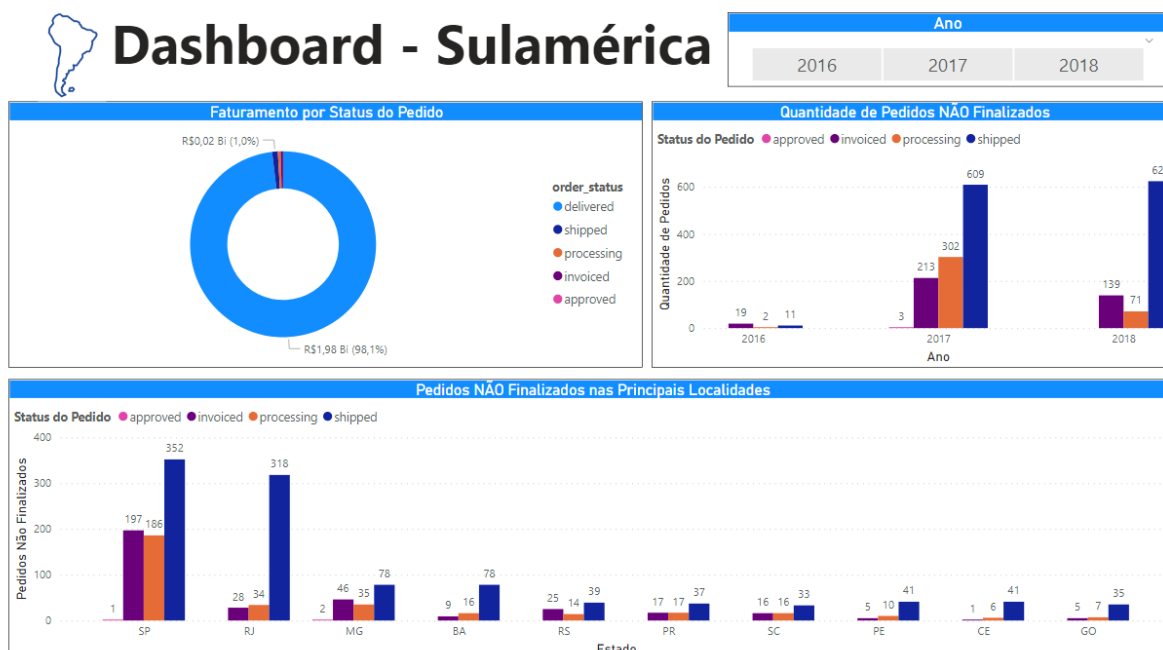


Figura 16 – Dashboard Loja Sulamérica – Visão Customer Success

7 Conclusão

Este projeto serviu como oportunidade para aplicar o processo de BI convencional para ajudar a Loja Sulamérica a ter o suporte necessário para análise dos dados de suas vendas para as devidas tomadas de decisões. Com a elaboração do modelo multidimensional e a respectiva criação do Data Warehouse, o cliente poderá inferir análises sobre o seu negócio no período verificado de maneira rápida e intuitiva, permitindo decisões rápidas e assertivas.

Com sua satisfação nos relatórios apresentados e nas análises possíveis de serem feitas, o cliente se mostrou interessado em expandir esta parceria, incluindo os dados mais recentes de ordem de vendas e a possibilidade de disponibilizar mais dados para que as análises possam ser feitas com baixa granularidade, permitindo o acompanhamento de maiores detalhes onde oportunidades possam ser verificadas para tomadas de ações mais direcionadas para o aumento das vendas.

8 Arquivos

8.1 pentaho_transformations.zip

Traz os arquivos de transformações feitos no Pentaho Data Integration para as cargas transacionais e multidimensionais.

8.2 scripts.zip

Traz todos os scripts utilizados no projeto em arquivos .sql contendo a construção da base transacional e da base multidimensional do projeto.

8.3 Modelos_power_architect_projeto.zip

Traz todos os modelos, transacionais e multidimensionais, desenvolvidos no software SQL Power Architect.

8.4. Archive.zip

Traz os arquivos em excel das bases de dados disponibilizadas pela empresa.

8.5 PBI-Projetov2.PBIX

Arquivo do Dashboard feito no Power BI.