**Forma

Descrição gerada automaticamente com confiança baixa**

**Loja Sulamérica**

**Projeto da Disciplina de BI**

Prof. Anderson Nascimentoprof.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, multidimensiononais 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](#_Toc87298655)

[2 Estudo de Caso 4](#_Toc87298656)

[2.1 Descrição do Estudo de Caso 4](#_Toc87298657)

[3 Descrição do Modelo Transacional 4](#_Toc87298658)

[3.1 Fonte 1 – Diagrama de modelo de dados físico. 4](#_Toc87298659)

[3.2 Fonte 2 – Entrevista. 5](#_Toc87298660)

[3.3 Descrição do projeto ETL Transacional 6](#_Toc87298661)

[4 Modelo Multidimensional 8](#_Toc87298662)

[4.1 Descrição 8](#_Toc87298663)

[5 Projeto de ETL 9](#_Toc87298664)

[5.1 Descrição do Projeto de ETL 9](#_Toc87298665)

[6 Dashboard 12](#_Toc87298666)

[6.2 Telas do Dashboard 12](#_Toc87298667)

[7 Conclusão 15](#_Toc87298668)

[8 Arquivos 15](#_Toc87298669)

[8.1 pentaho\_transformations.zip 15](#_Toc87298670)

[8.2 scripts.zip 15](#_Toc87298671)

[8.3 Modelos\_power\_achitect\_projeto.zip 15](#_Toc87298672)

[8.4. Archive.zip 15](#_Toc87298673)

[8.5 PBI-Projetov2.PBIX 15](#_Toc87298674)

# 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.

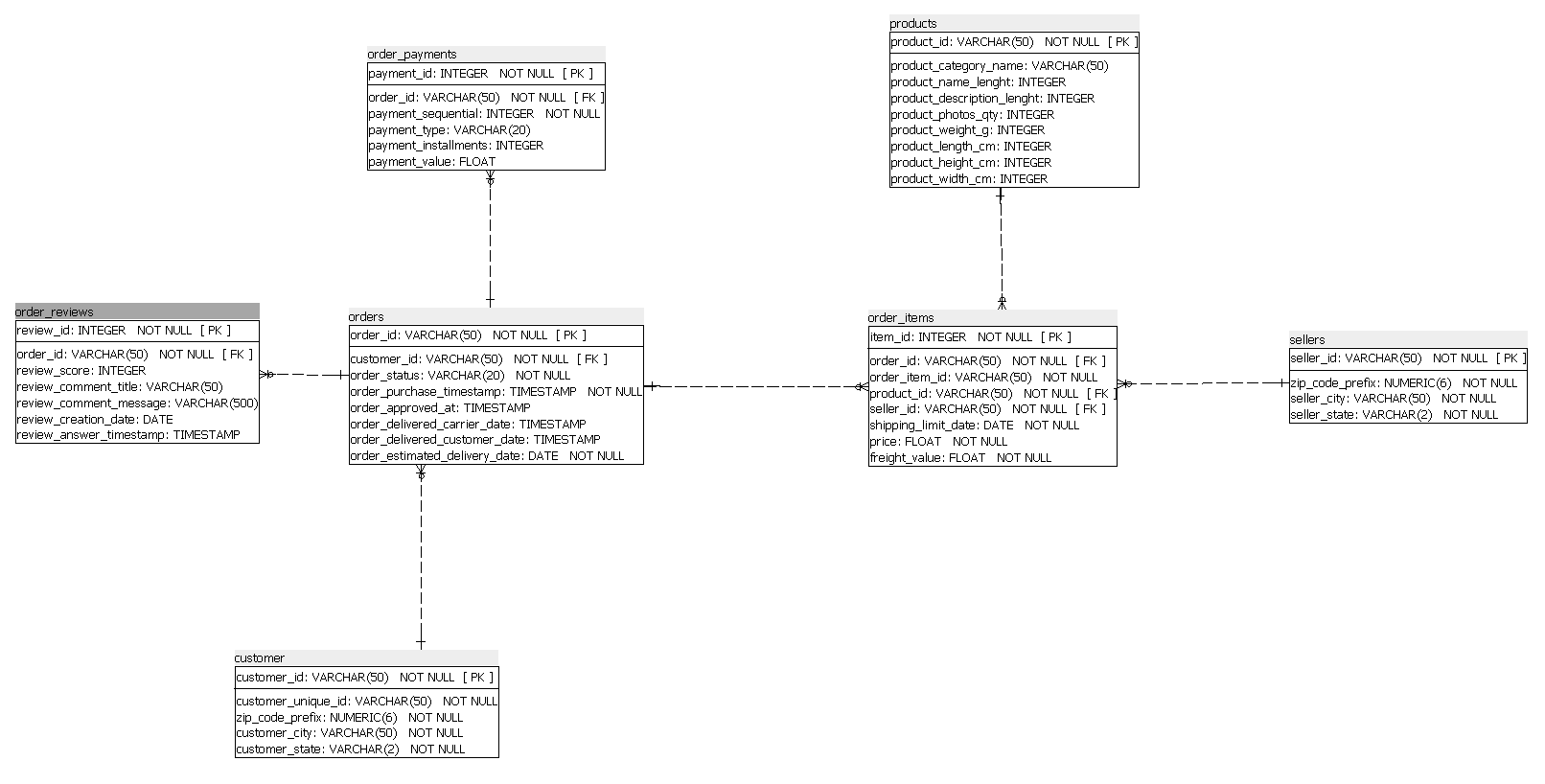


Figura 1 – Diagrama do modelo físico do Sistema de ordem de pedidos da Loja Sulamérica

## 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)

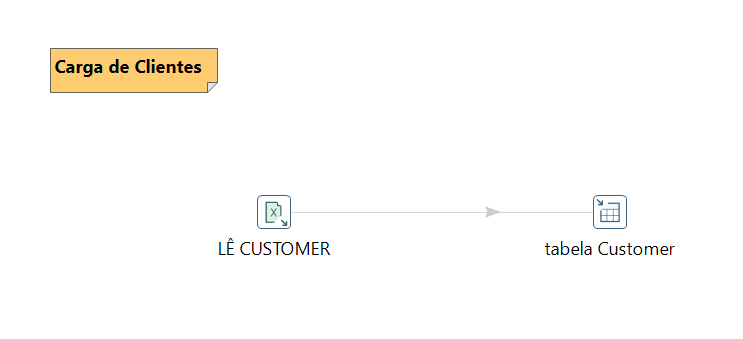


Figura 2 – Carga transacional customers

1. Vendedores (carga\_transacional\_sellers.ktr)

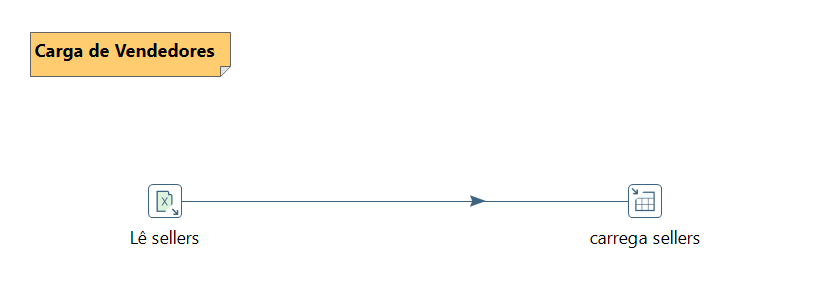


Figura 3 – Carga transacional sellers

1. Produtos (carga\_transacional\_products.ktr)

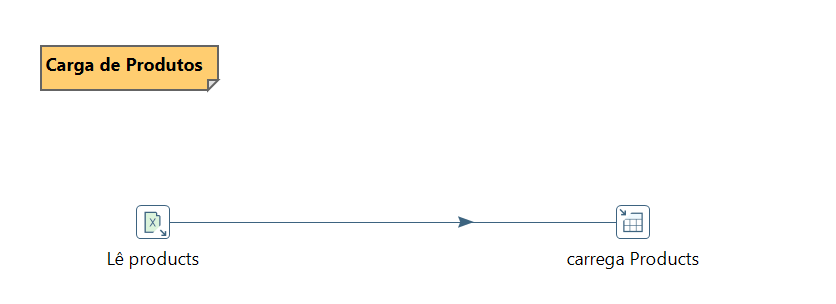


Figura 4 – Carga transacional products

1. Pedidos (carga\_transacional\_orders.ktr)

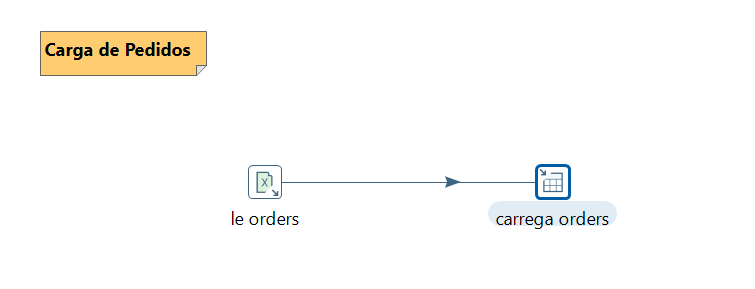


Figura 5 – Carga transacional orders

1. Itens de pedidos (carga\_transacional\_orderitens.ktr)

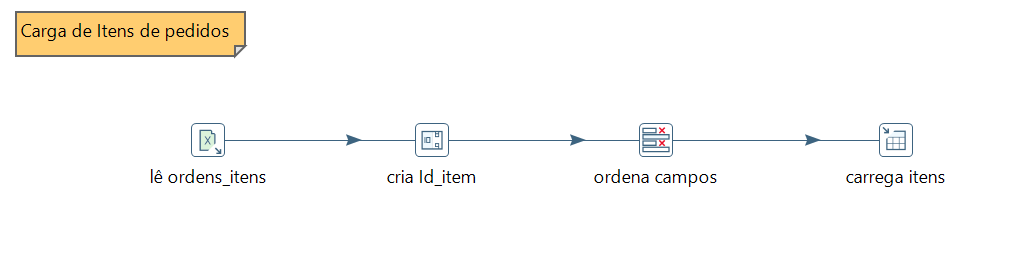


Figura 5 – Carga transacional orders\_itens

1. Ordens de pagamentos (carga\_transacional\_orderpayments.ktr)

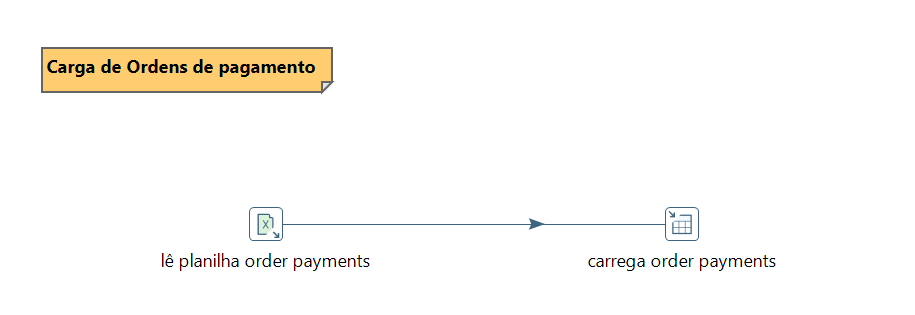


Figura 6 – Carga transacional order\_payments

1. Avaliações (carga\_transacional\_orderreviews.ktr)

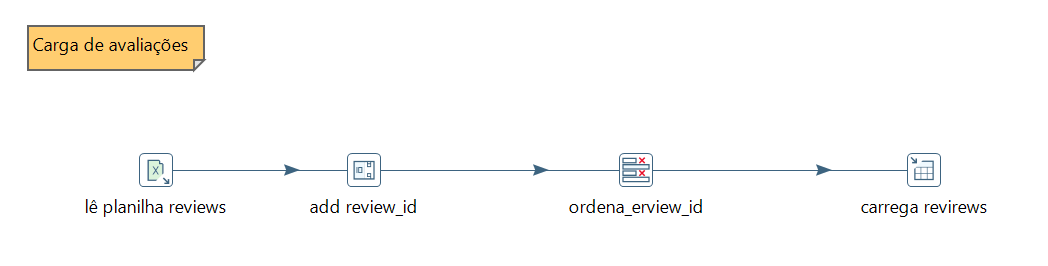


Figura 7 – Carga transacional order\_reviews

# 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.

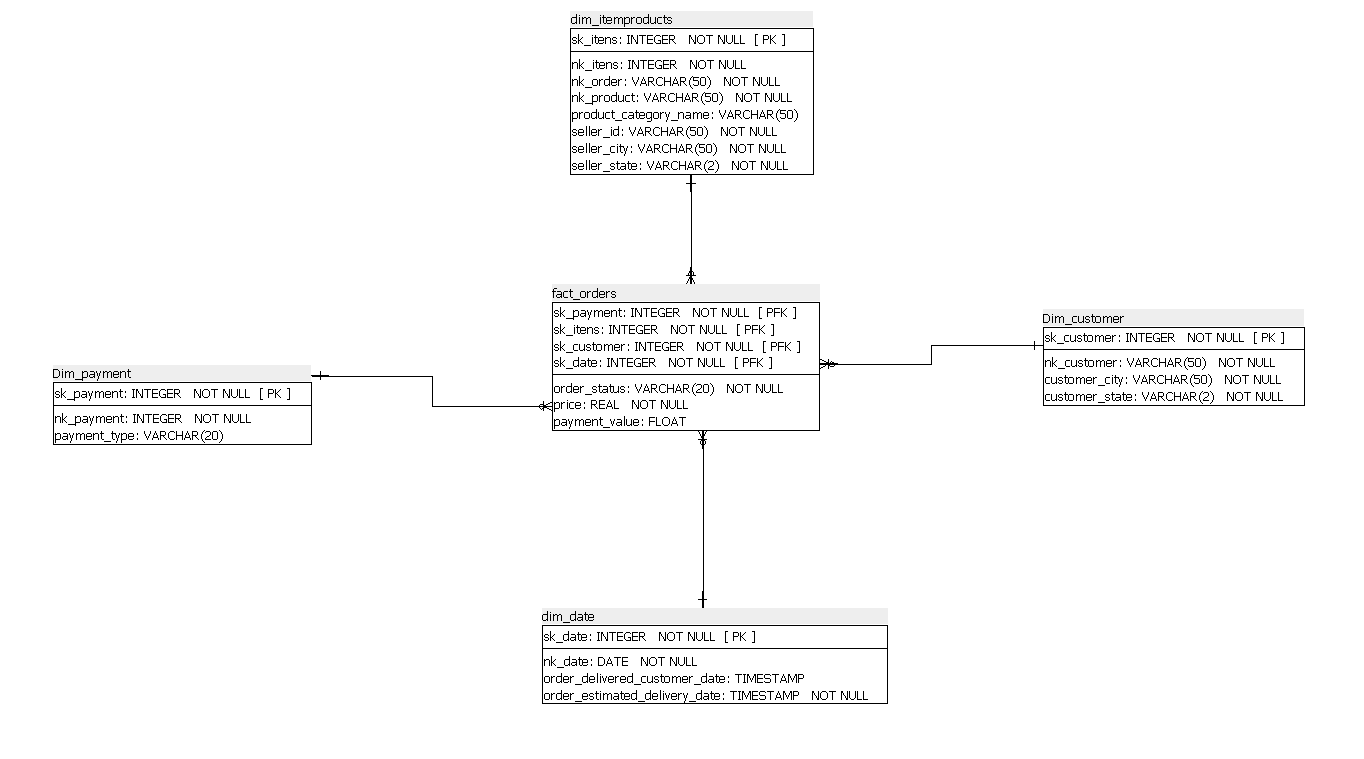


Figura 8 – Modelo Multidimensional

# 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:

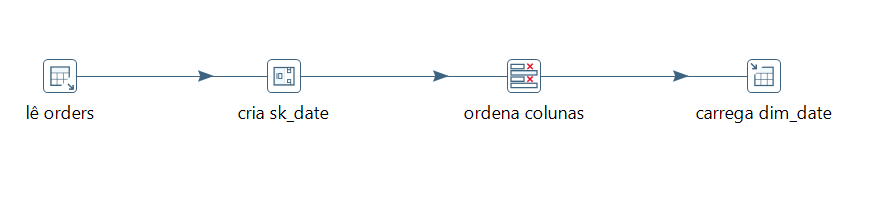


Figura 9 – ETL dim\_date

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 **R**io **D**e **J**aneiro, passa agora a ser transformado para Rio **d**e 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:

Gráfico

Descrição gerada automaticamente com confiança média

Figura 11 – ETL dim\_itemsproducts

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 utilizado 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:

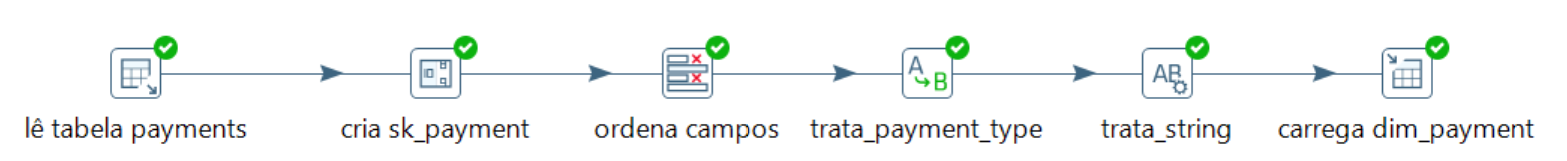


Figura 12 – ETL dim\_payments

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:

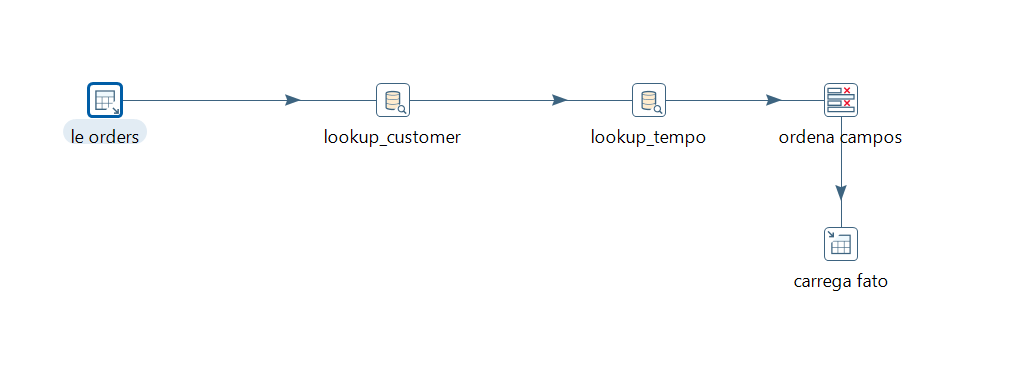


Figura 13 – ETL dim\_itemsproducts

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 “lê 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.

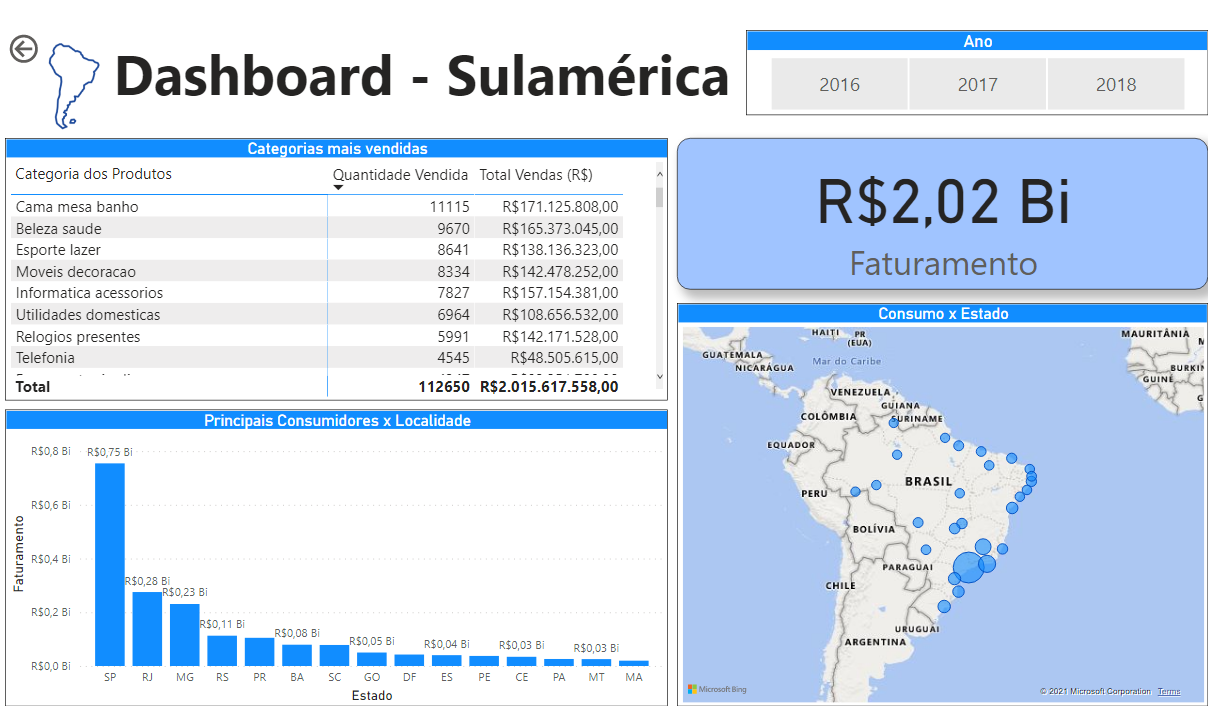


Figura 14 – Dashboard Loja Sulamérica – Visão Geral

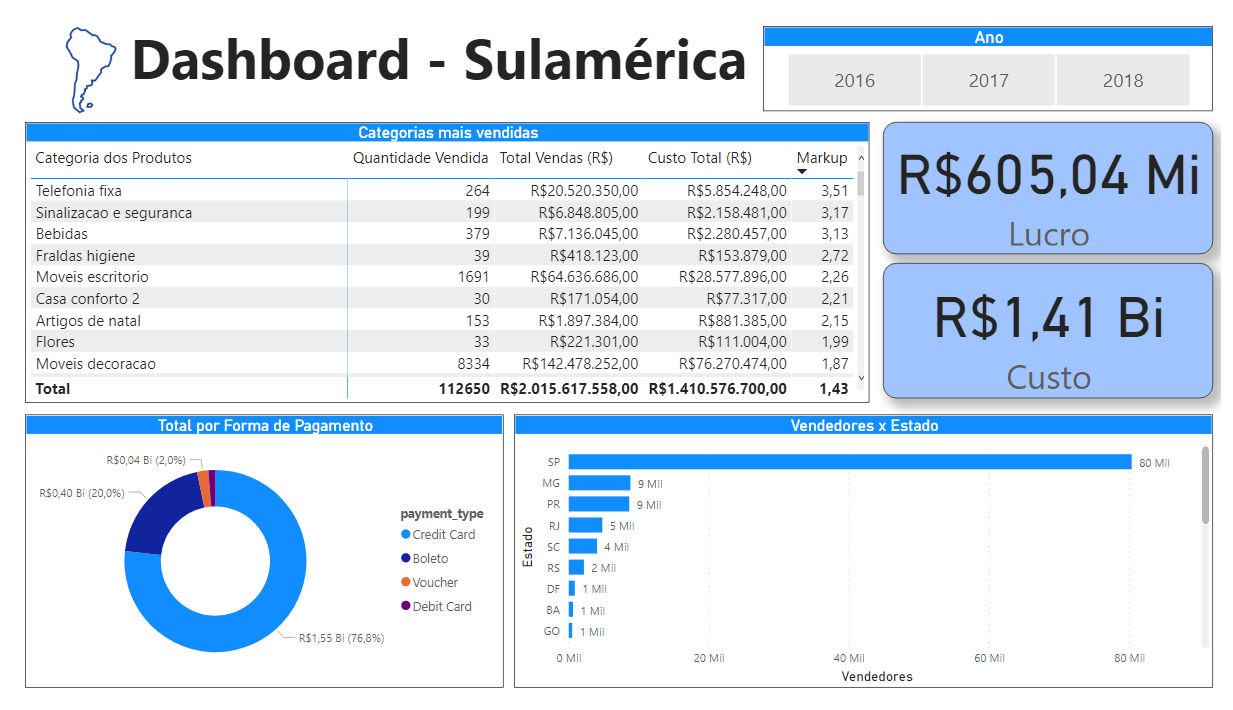


Figura 15 – Dashboard Loja Sulamérica – Visão Financeiro

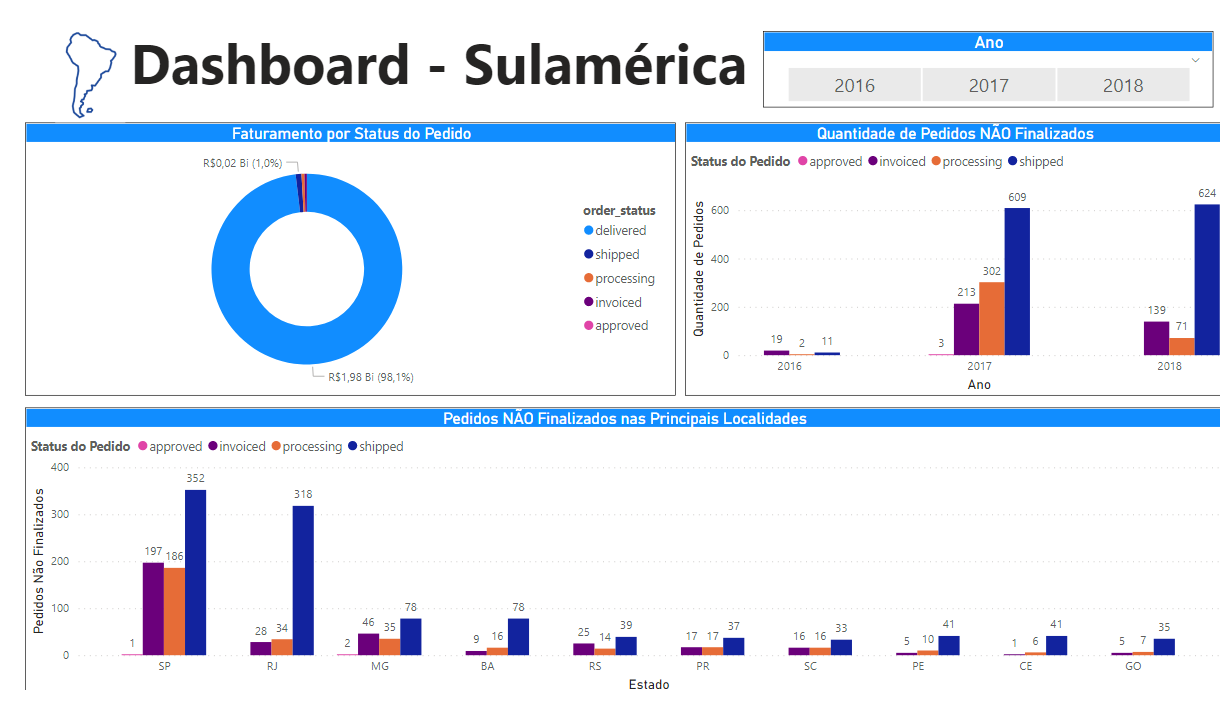


Figura 16 – Dashboard Loja Sulamérica – Visão Customer Success

# 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\_achitect\_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.