

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**NÚCLEO DE EDUCAÇÃO A DISTÂNCIA

Pós-graduação *Lato Sensu* em Gestão e Análise Estratégica de Dados

**RELATÓRIO TÉCNICO**

MATHEUS BUNIOTTO AGUILAR

Belo Horizonte

2021

**SUMÁRIO**

[1. Introdução 3](#_Toc87695275)

[1.1. Contexto 3](#_Toc87695276)

[1.2. Objetivos 3](#_Toc87695277)

[1.3. Público alvo 3](#_Toc87695278)

[2. Modelos de Dados 4](#_Toc87695279)

[2.1. Modelo Dimensional 4](#_Toc87695280)

[2.2. Fatos e Dimensões 4](#_Toc87695281)

[3. Integração, Tratamento e Carga de Dados 5](#_Toc87695282)

[3.1. Fontes de Dados 5](#_Toc87695283)

[3.2. Processos de Integração e Carga (ETL) 6](#_Toc87695284)

[4. Camada de Apresentação 8](#_Toc87695285)

[4.1. Dashboard 8](#_Toc87695286)

[5. Registros de Homologação 10](#_Toc87695287)

[6. Conclusões e Recomendações 12](#_Toc87695288)

[6.1. E-commerce no Brasil 12](#_Toc87695289)

[6.2. Análise exploratória dos dados 13](#_Toc87695290)

[7. Links 16](#_Toc87695291)

[REFERÊNCIAS 16](#_Toc87695292)

# 1. Introdução

## 1.1. Contexto

O comércio eletrônico (e-commerce) tem ganhando cada vez mais espaço na vida dos consumidores. Com o aumento do consumo e acesso a esse tipo de comércio surgem novos desafios, por exemplo, a enorme quantidade de dados geradas a cada transação se torna um ativo de grande valia para as empresas, porém elas carecem de análises ágeis e efetivas que transformem os dados em informações para a empresa. Devido ao grande volume de dados, se torna cada vez mais difícil para as pessoas da organização extrair informações de forma eficiente.

## 1.2. Objetivos

Este projeto tem como objetivo extrair informações através de ferramentas de visualização de dados para que os stakeholders do projeto possam analisar e tomar decisões embasadas, melhorando a eficácia e maximizando os lucros da empresa. Além disso, o projeto visa cruzar informações com a utilização de técnicas de mineração e processamento de dados para que possamos extrair informações relevantes do conjunto de dados.

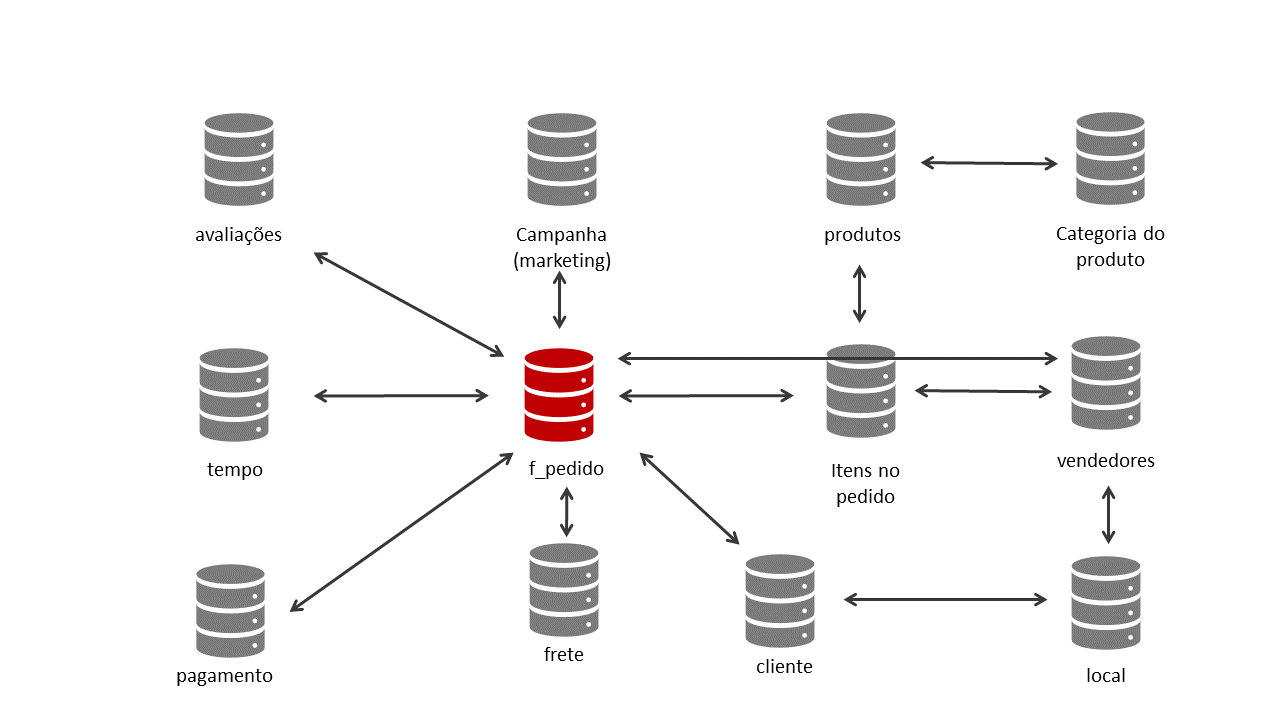
## 1.3. Público alvo

Os dashboards e análises desenvolvidas para o projeto tem como público alvo os integrantes da organização responsáveis pela tomada de decisão. Por se tratar de um e-commerce, esses usuários já possuem uma experiência maior com tecnologia e a utilizam durante seu dia-a-dia para aumentar sua produtividade. Além disso, os stakeholders desse projeto tem uma visão analítica do negócio, buscando novas oportunidades para aumentar a lucratividade da empresa.

# 2. Modelos de Dados

## 2.1. Modelo Dimensional

O conjunto de dados escolhido para este projeto é do marketplace Olist, disponível no Kaggle¹, que reúne diversos conjuntos de dados públicos utilizados para estudos e projetos.



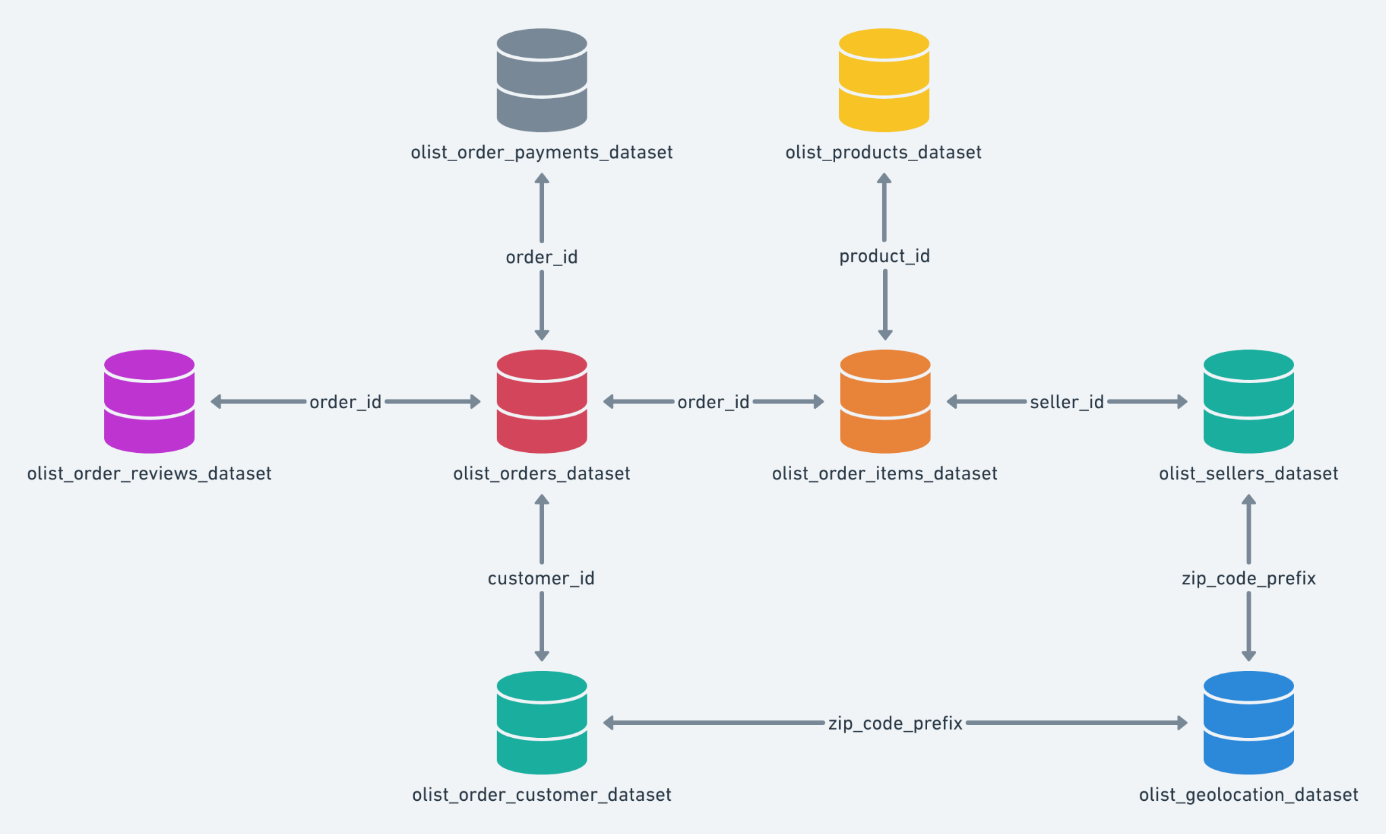
## 2.2. Fatos e Dimensões

A tabela “fato” definida para esse projeto será o pedido feito no marketplace. Já as tabelas de dimensões para análise serão: clientes, pagamento, frete, local, tempo(calendário), avaliações, produtos, itens no pedido, categoria do produto e campanha de marketing.

# 3. Integração, Tratamento e Carga de Dados

## 3.1. Fontes de Dados

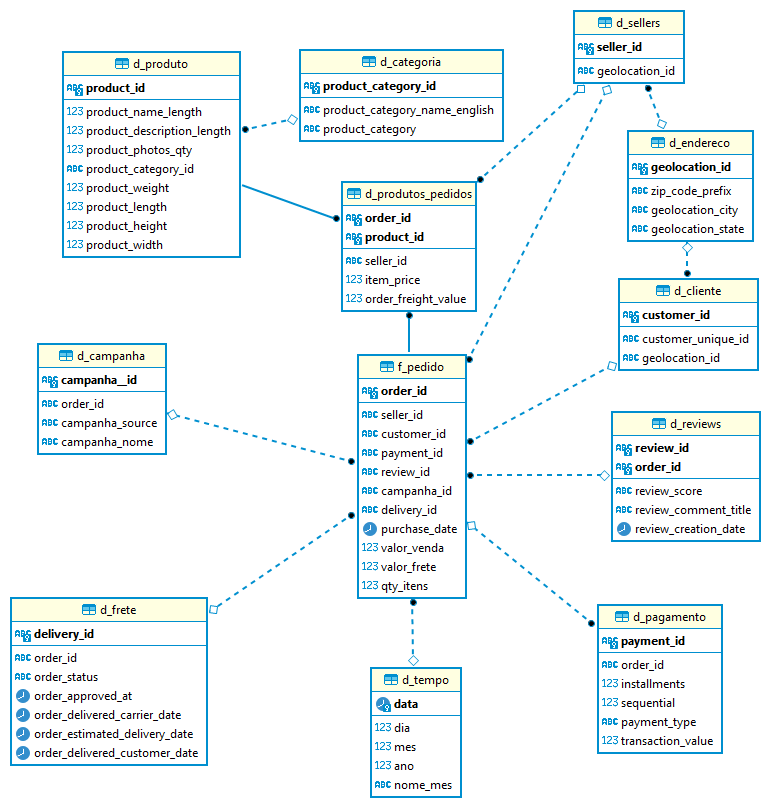
Os dados da Olist disponíveis para download originalmente não possuem uma tabela fato específica, porém, a tabela de pedidos centraliza a maioria das informações, sendo a chave order\_id a conexão entre a maioria das dimensões.

**Figura 1** -Diagrama do diagrama modelo dimensional original

**Fonte:** Kaggle¹

Para esse projeto, os dados foram enriquecidos como uma fonte de dados externa fictícia que indica a fonte de tráfego que gerou o fato pedido no banco de dados. Além disso, será utilizada uma dimensão calendário para relacionamento das datas entre as tabelas. Também optei por separar a tabela frete, criando uma nova tabela de dimensão, diferente do modelo original.

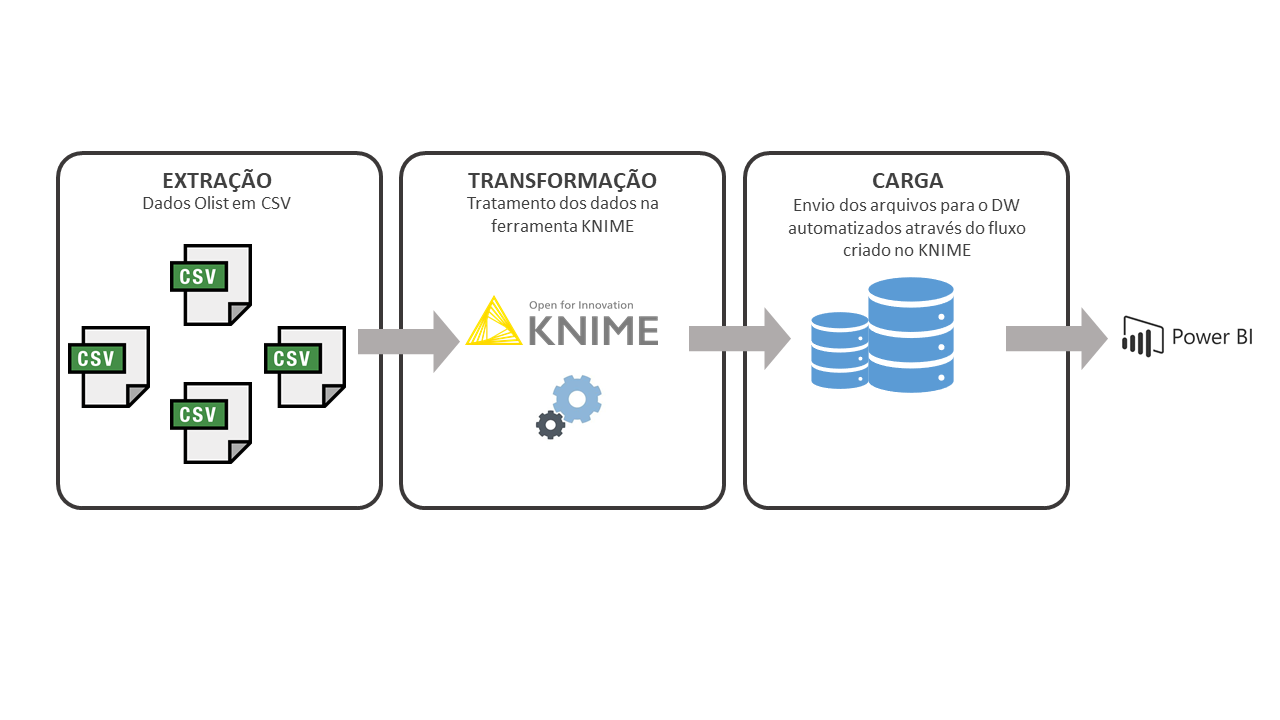
**Figura 2** - Modelo dimensional do projeto



**Fonte:** Elaborado pelo autor

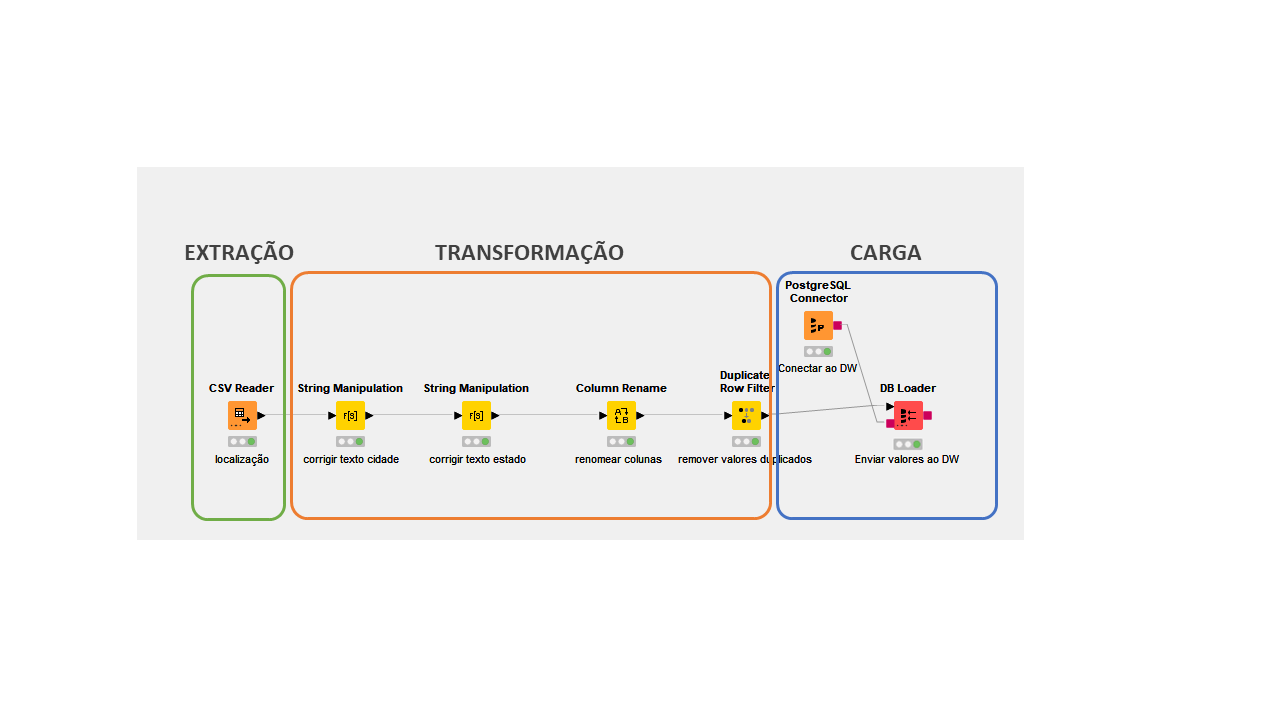
## 3.2. Processos de Integração e Carga (ETL)

Para o processo de ETL utilizarei o software KNIME, que permite criar fluxos de tratamento, processamento e carga dos dados em um banco de dados ou planilha. Após isso, os dados serão utilizados no software Power BI para criação dos dashboards.

**Figura 3** - Fluxo ETL

**Fonte:** Elaborado pelo autor

Dentro do software KNIME podemos criar fluxos que permitem o tratamento e carga dos dados ao banco de dados de forma automatizada. Para esse projeto, foram utilizados os arquivos CSV disponibilizados através do Kaggle. Esses arquivos CSV são carregados no KNIME, tratados e enviados ao banco ou disponibilizados em formato CSV. Em uma aplicação real, os dados poderiam ser extraídos diretamente da fonte transacional, tratados e enviados ao banco de dados de forma automatizada.

**Figura 4** - Fluxo ETL KNIME

**Fonte:** Elaborado pelo autor

# 4. [Camada de Apresentação](https://docs.google.com/document/d/1LvfrykIc0e6s53G0qomOarDeq-5A6h290mgEuqAuBMs/edit#heading=h.17dp8vu)

## 4.1. Dashboard

Para criação do Dashboard foi utilizado o *software* PowerBI, onde foram criados 3 *dashboards* diferentes. O primeiro, de nível estratégico, resume as principais métricas e KPIs do negócio ao longo do tempo e por região. Aqui, temos a visão mais abrangente do fato venda e a consolidação dos valores mais importantes para a rentabilidade do negócio.

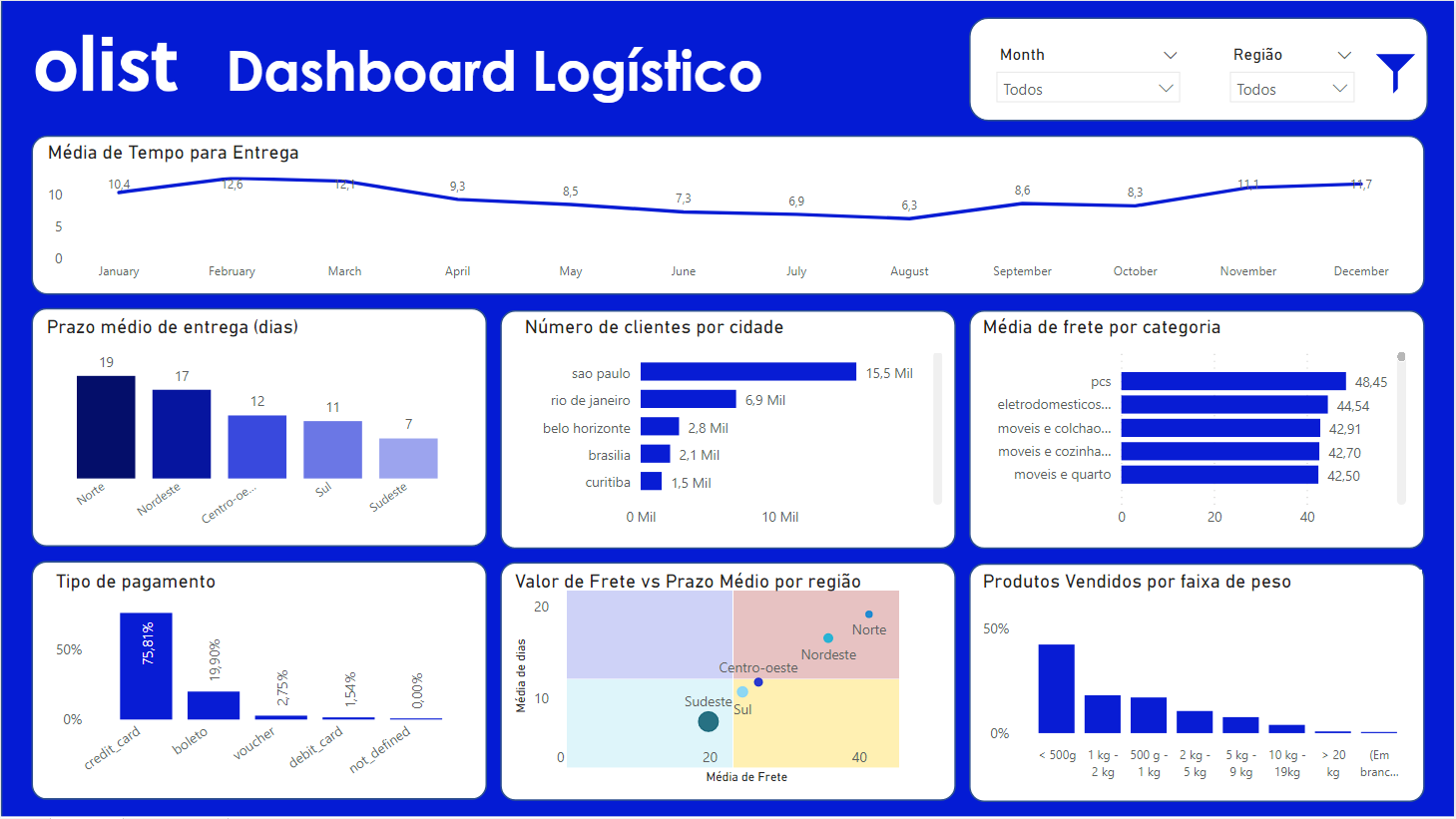
Para o *dashboard* de visão estratégica foram definidos os KPIS: receita, número de pedidos e ticket médio (calculado através da receita total divida pelo número de pedidos). Para visualização destes indicadores foi utilizado gráficos de linha para medir evolução ao longo do tempo. Nesse dashboard também temos informações de receita por estado, com a visão de mapa e bolhas, além de uma análise de pedidos e inventário por categoria.

**Figura 5** – Dashboard com visão estratégica

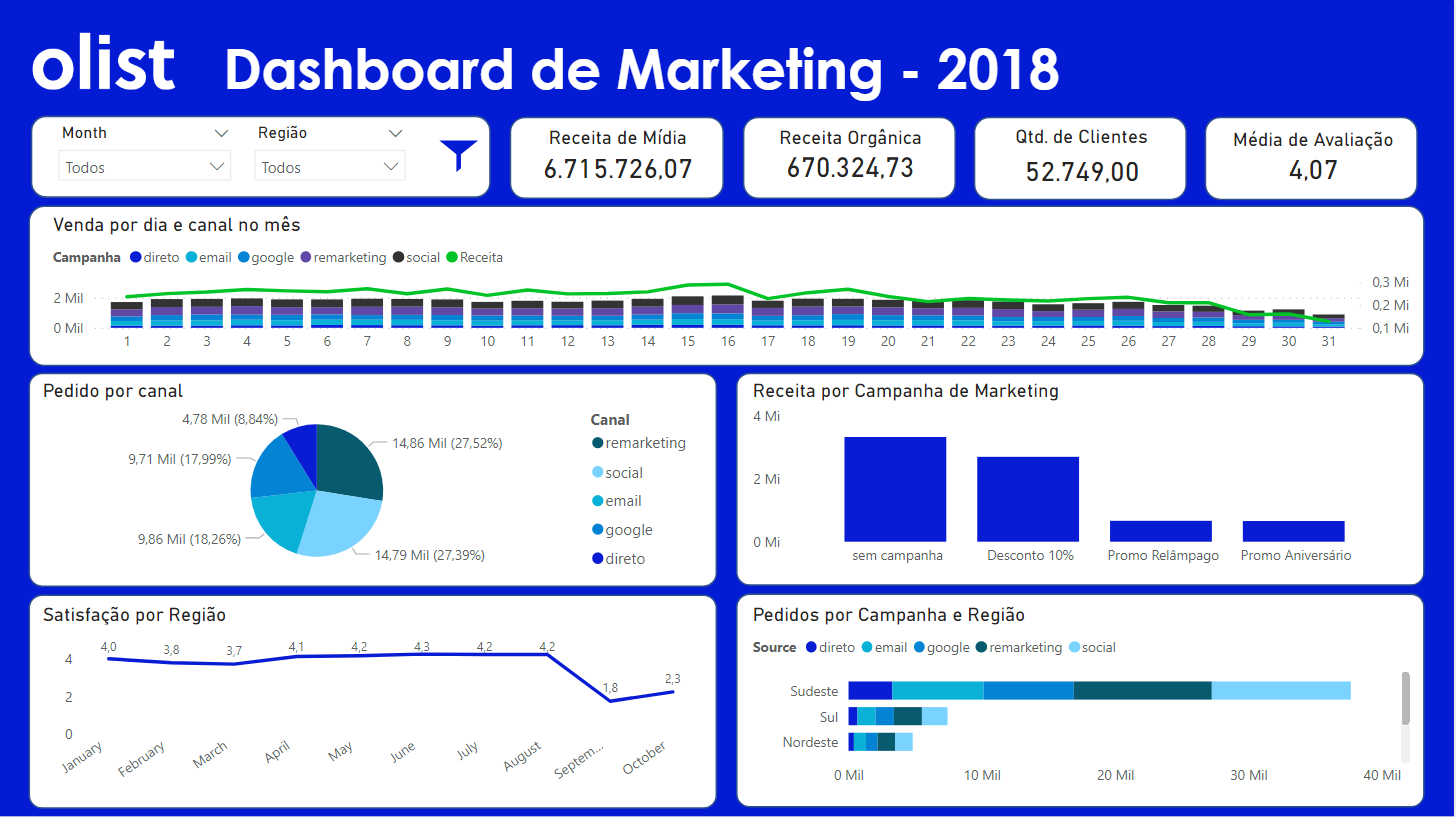
**Fonte**: elaborada pelo autor com *o software* PowerBI.

Além da visão de nível estratégico, também foram criados outros 2 *dashboards* com as visões táticas e operacionais do negócio. Para isso, o fato é analisado com maior detalhamento para entender a venda realizada com grau de granularidade menor.

**Figura 6** – Dashboard com visão tática detalhando as vendas com análise logística.

**Fonte**: elaborada pelo autor com *o software* PowerBI.

No Dashboard com visão logística (Figura 6) é possível analisar o valor do frete, prazo de entrega e como essas 2 dimensões da venda se relacionam com as dimensões de tempo, região e categoria do produto. O gráfico de linha apresenta a média do frete ao longo dos meses para que os stakeholders possam identificar possíveis períodos problemáticos. Além disso, foram utilizados gráficos de barras (horizontais e verticais) para detalhamento do prazo médio de frete por estado, número de clientes por cidade, valor médio do frete por categoria de produto, tipo de pagamento e a faixa de peso dos produtos vendidos. Nessa visão, também temos um quadrante que identifica a relação entre valor do frete, prazo de entrega e número de pedidos, nesse gráfico podemos ver quais os estados possuem melhor relação entre preço e prazo de entrega gerando oportunidades para melhorias.

**Figura 7** – Dashboard com visão operacional para equipe de marketing.

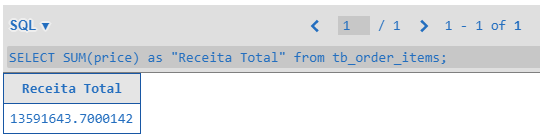
**Fonte**: elaborada pelo autor com *o software* PowerBI.

No Dashboard de Marketing (Figura 7), temos o fato venda detalhado por canal, campanha de marketing e região. Além disso, também é possível acompanhar métricas relacionadas ao cliente, como por exemplo a nota de avalição fornecida pelo cliente ao fato venda. Por se tratar de um dashboard com objetivo operacional, podemos ver a evolução das métricas ao longo dos dias e a tendência de queda ou crescimento com o decorrer do tempo.

# 5. Registros de Homologação

Para homologação dos resultados exibidos nos dashboards, foram realizadas consultas em SQL feitas contra a fonte de dados original e pelo confrontamento do que é exibido como resultado nos *dashboards*.

Nas figuras abaixo podemos visualizar os resultados em SQL e o respectivo resultado apresentado para a segmentação semelhante no dashboard.

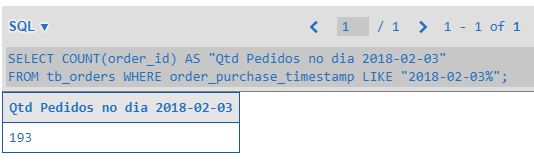
**Figura 8** – Homologação de receita total



**Fonte**: elaborada pelo autor com *o software* PowerBI.

**Fonte**: elaborada pelo autor com *o software* PowerBI.

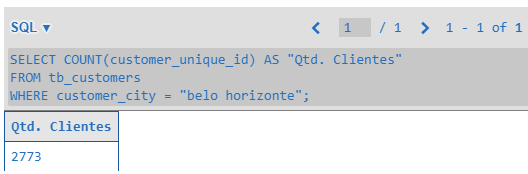
**Figura 9** – Homologação de pedidos com filtro de data

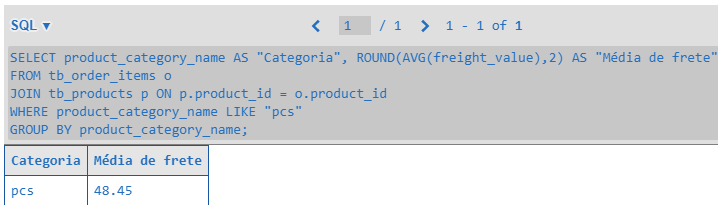


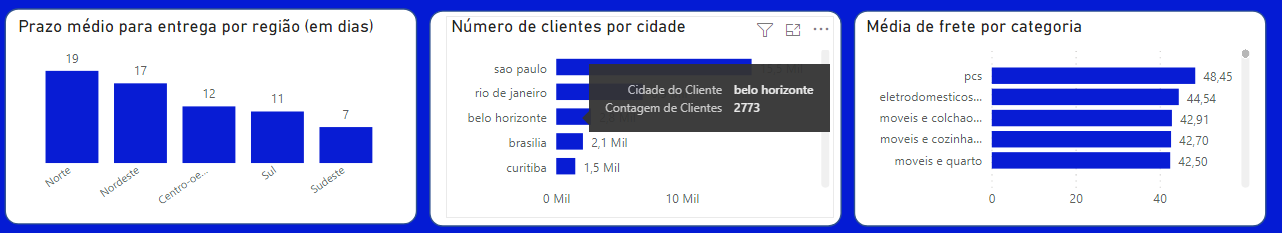


**Fonte**: elaborada pelo autor com *o software* PowerBI.

**Figura 10** – Homologação de frete e clientes

****

****

****

**Fonte**: elaborada pelo autor com *o software* PowerBI.

# 6. Conclusões e Recomendações

## 6.1. E-commerce no Brasil

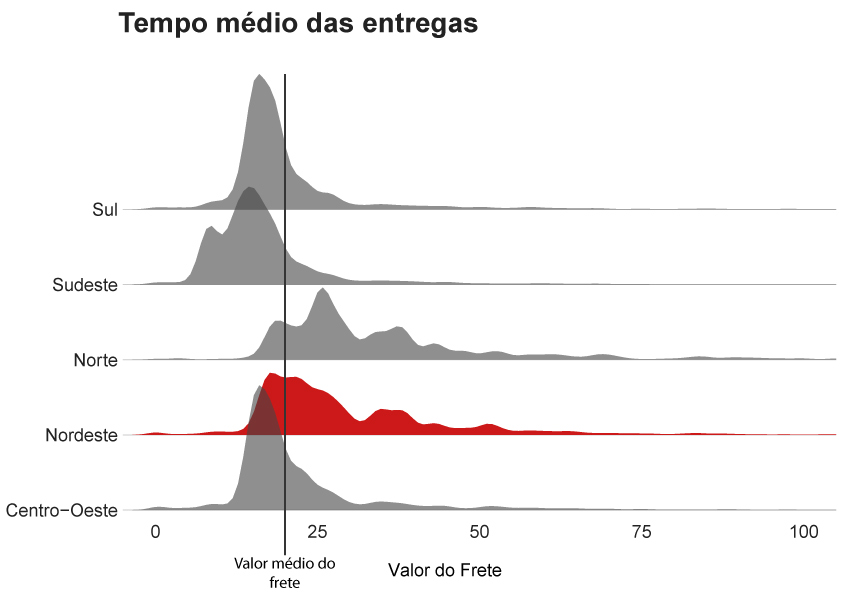
Apesar da predominância das regiões Sul e Sudeste como os principais mercados de e-commerce no Brasil, as regiões Norte e principalmente a região Nordeste apresentam um enorme crescimento e potencial no mercado nacional. É o que nos indica a 44ª edição do Webshoppers, um dos mais amplos relatórios de sobre e-commerce no Brasil, elaborado pela Ebit | Nielsen.

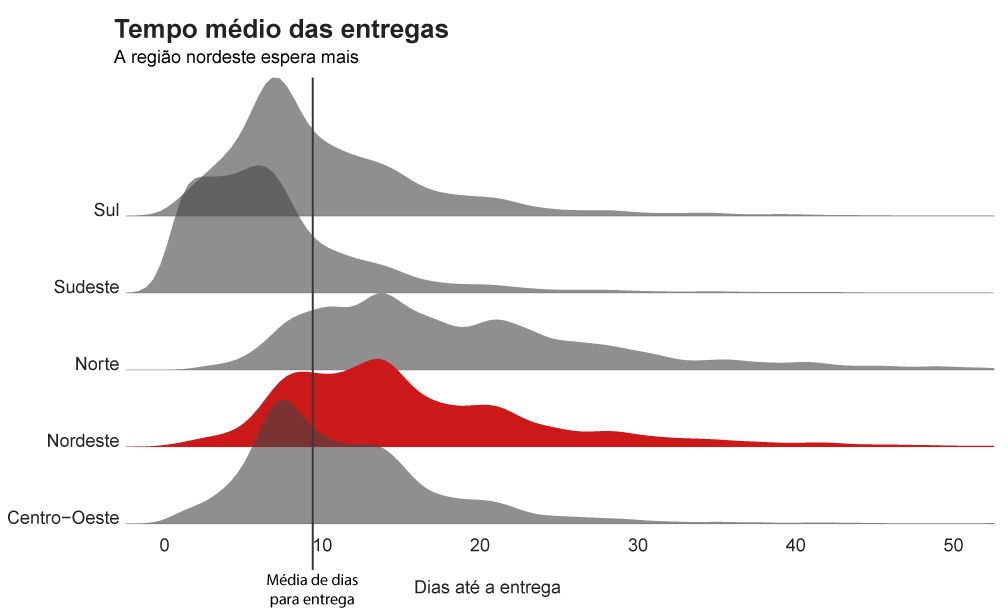
Os dados MCC-ENET também apontam um crescimento acumulado de mais de 50% no Índice de Vendas Online nos últimos 12 meses. Outro ponto importante apontado pelos relatórios do Webshoppers é a importância do valor do frete na satisfação e preferências dos consumidores online.

A partir desse cenário apresentado vamos analisar os dados abertos da empresa Olist para entender essas relações a partir do ponto de vista de frete, regiões e satisfação do consumidor.

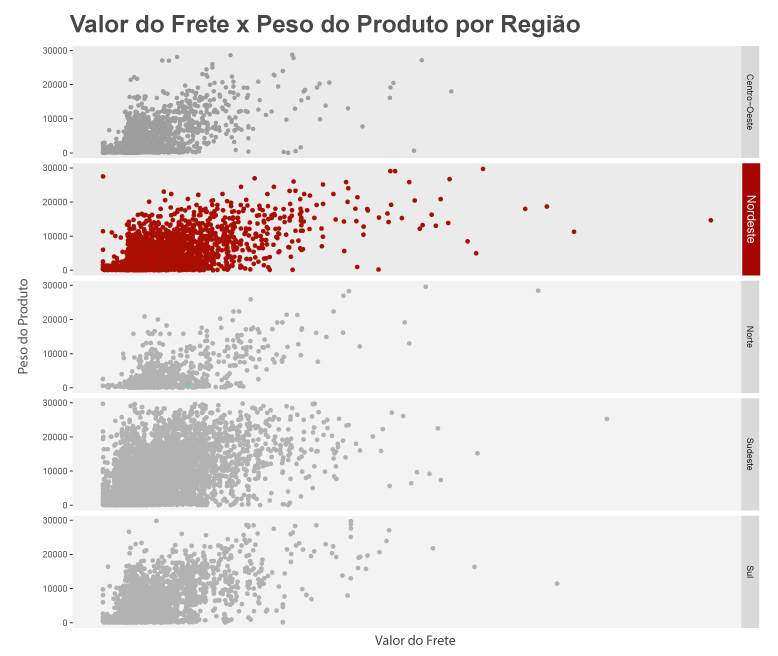
## 6.2. Análise exploratória dos dados

Os clientes da região norte e nordestes são os que pagam mais pelo frete nos pedidos na Olist. Podemos ver que a maioria dos clientes dessas regiões pagam valores acima do valor médio do frete cobrado no marketplace.

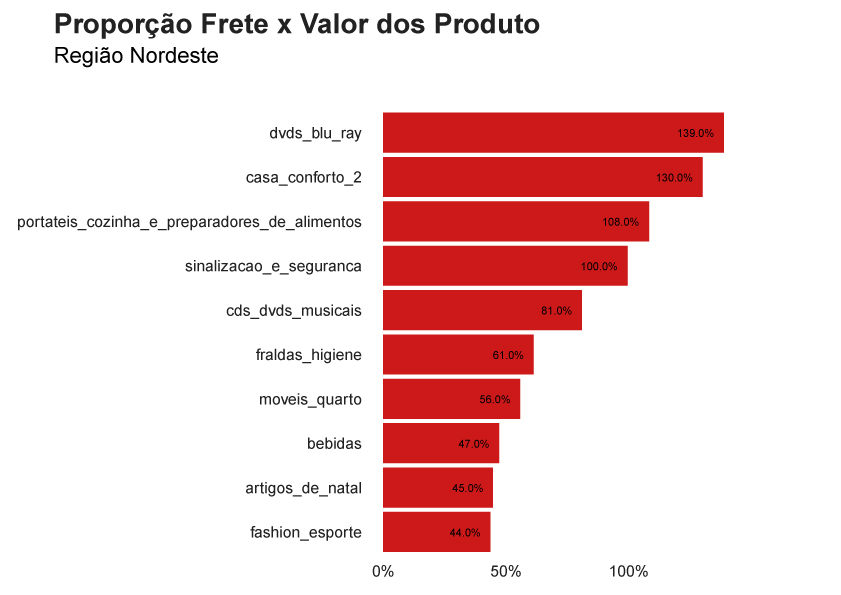


Além disso, esses clientes também são os que mais esperam para receber os produtos que foram compradores.

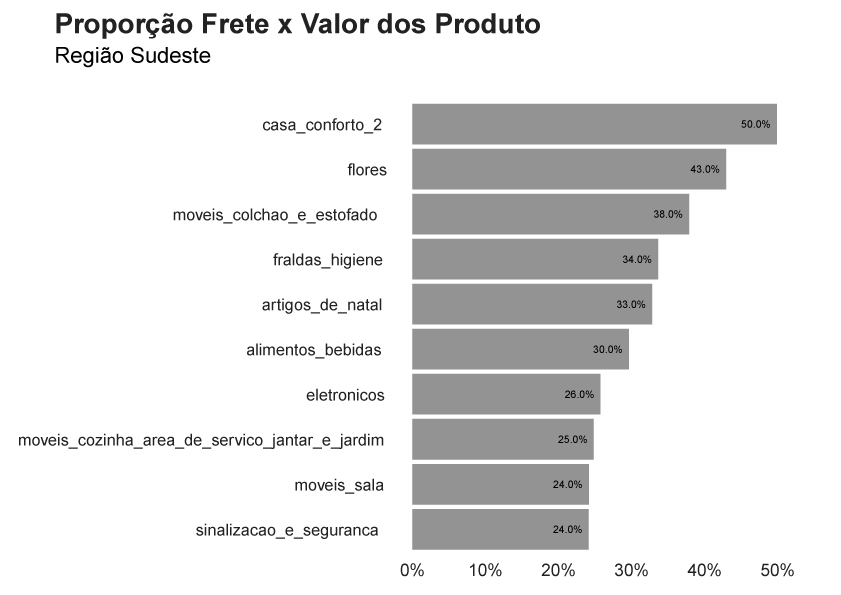
Vale ressaltar que esse valor maior não é influenciado por pedidos com peso ou volume maiores do que a média para a região nordeste, como podemos ver na gráfico de distribuição entre peso e valor do frete.



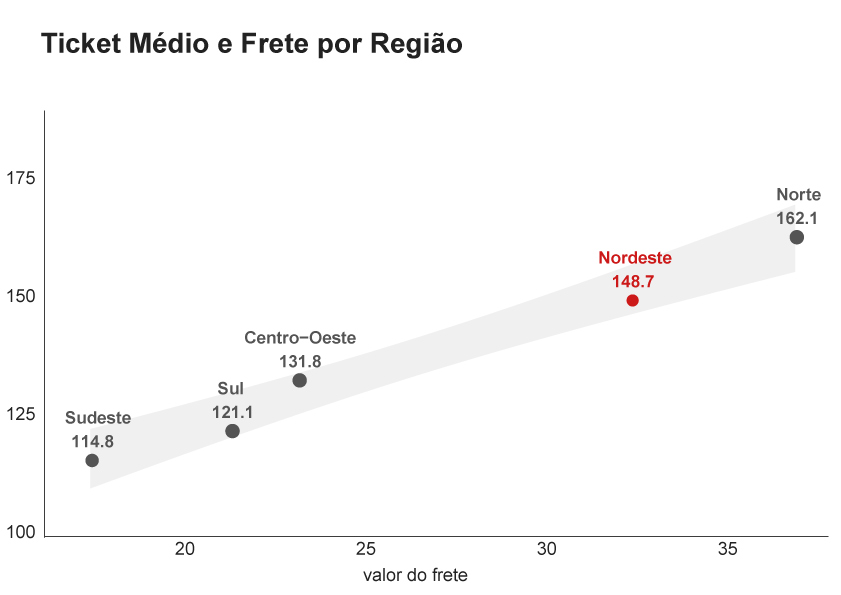
Para os clientes da região nordeste, o valor do frete ultrapassa 100% em relação ao valor do produto comprado para algumas categorias, como por exemplo Casa e Conforto (2), nessa categoria, em média, o valor do frete representa 139% o valor do produto comprado.



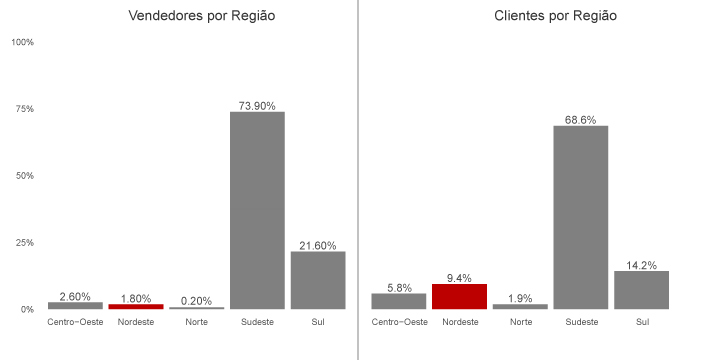
Quando comparamos com os clientes da região Sudeste, esse valor não ultrapassa 50% do valor pago pelo produto.



Esse valor mais alto do frete em relação aos produtos faz com que o Ticket Médio para essas regiões seja maior, indicando que o cliente faz essa comparação de frete e valor do produto no momento da compra, tendo preferência por produtos com valores maiores, compensando o alto valor do frete que será pago por ele.

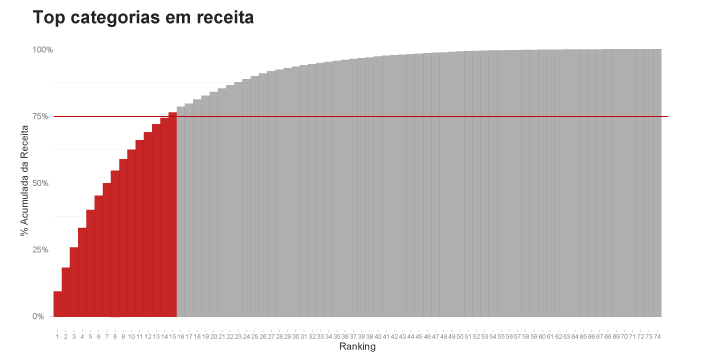


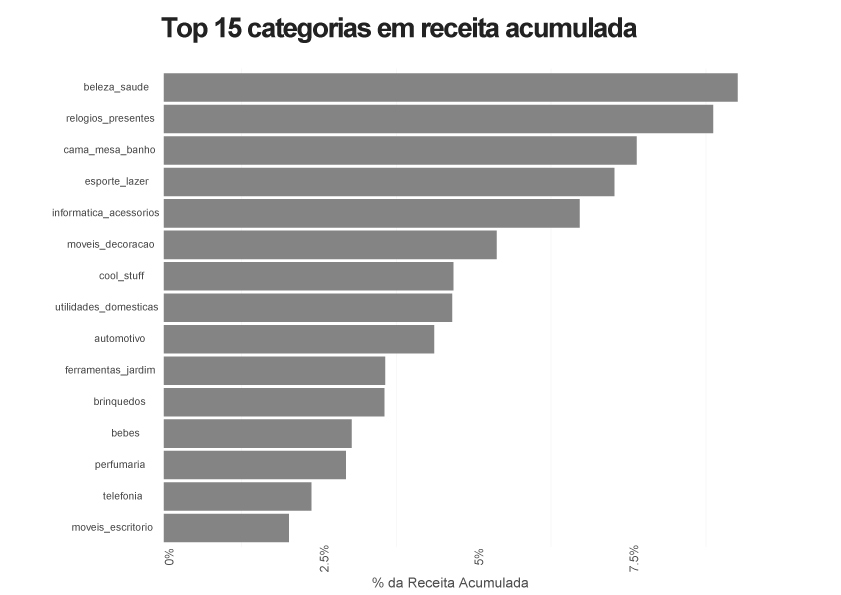
Um dos fatores que pode contribuir para esse alto valor no frete é a quantidade de vendedores presentes na região nordeste, esses vendedores representam apenas 1,80% do total de vendedores que vendem no marketplace. Sendo assim, a maioria dos produtos comprados pela região nordeste são enviados de outras regiões. Por outro lado, o número de compradores da região nordeste representa 9,4% do total, sendo a terceira maior região em número de compradores.



Uma estratégia de captação de novos vendedores da região nordeste poderia ser implementada visando aumentar o inventário de produtos disponíveis para envio partindo dessa região, oferecendo um frete mais atraente para os compradores regionais e dessa forma aumentar a participação de mercado da Olist.

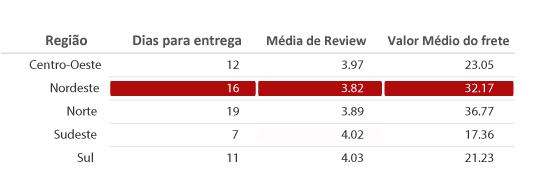
Além disso, pode ser oferecido subsídios de frete para categorias estratégicas, reduzindo o valor do frete para a região nordeste, independentemente da região do vendedor. Apenas 15 categorias de produtos representam mais de 75% do total da receita acumulada pelo marketplace, essas categorias poderiam ser o ponto de partida para implementação de estratégias de redução de frete para os compradores da região nordeste.



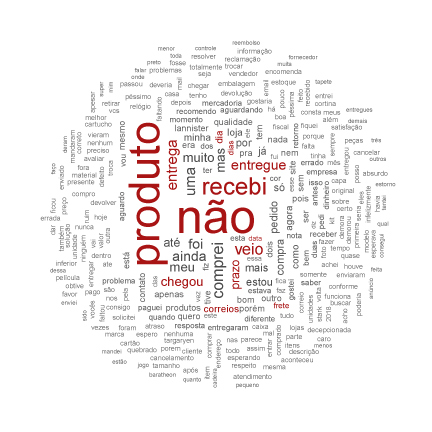


Uma estratégia focada em oferecer benefícios e condições melhores de frete para categorias e produtos com grande volume de demanda poderá trazer benefícios perceptíveis logo de início, validando ou desvalidando a hipótese inicial.

Outro fator que é impactado pelo prazo e valor do frete é a avalição do cliente em relação ao pedido, como podemos ver através da tabela abaixo, as regiões com maior prazo e maior valor de frete possuem uma avalição média menor do que as outras.



Filtrando as avalições com valor abaixo de 3 em uma escala de 0 a 5, podemos analisar as palavras mais citadas nos comentários de avaliações e dessa forma perceber que o frete e termos relacionados a ele são de fato muito presentes nessas avaliações negativas.



Dessa forma, uma estratégia focada em melhores condições de frete para regiões, promovendo melhores condições de frete e por consequência uma maior penetração de mercado e melhor satisfação do cliente, se prova uma das estratégias válidas para escalar o negócio.

# 7. Links

Artefatos:

[Arquivos](https://drive.google.com/drive/folders/1j1tZJFqMAZeMcYzZUEXOQpOa0EjviY7n?usp=sharing): <https://drive.google.com/drive/folders/1j1tZJFqMAZeMcYzZUEXOQpOa0EjviY7n?usp=sharing>  
Vídeo de apresentação: [https:/youtu.be/KsRlvdCch8o](https://youtu.be/KsRlvdCch8o)

# REFERÊNCIAS

¹ KAGGLE. Conjunto de dados públicos de comércio eletrônico brasileiro por Olist. Disponível em: https://www.kaggle.com/olistbr/brazilian-ecommerce. Acesso em: 13 jul. 2021.

ALVARENGA, M.V. **Visualização de dados e sua importância para a tomada de decisão: uma aplicação usando o conjunto de dados de e-commerce da Olist.** Ouro Preto, UFOP, 2021.

ALBERTIN, A. L. **Comércio Eletrônico:** Modelo, Aspectos e Contribuições de sua Aplicação. 7 ed. São Paulo: Atlas, 2016

EBIT - **Pesquisa aplicativos de entrega no painel de consumidores online** Ebit, 2021.