

BRAZILIAN E-COMMERCE PUBLIC DATASET BY OLIST

Filipe Leonardo de Sousa
Gabriel Marques F Leal
Maria Paula Pinto de Andrade
Matheus Fonseca Chaves
Mileno Alexandre Barbosa Epifânio
Mille de Amorim

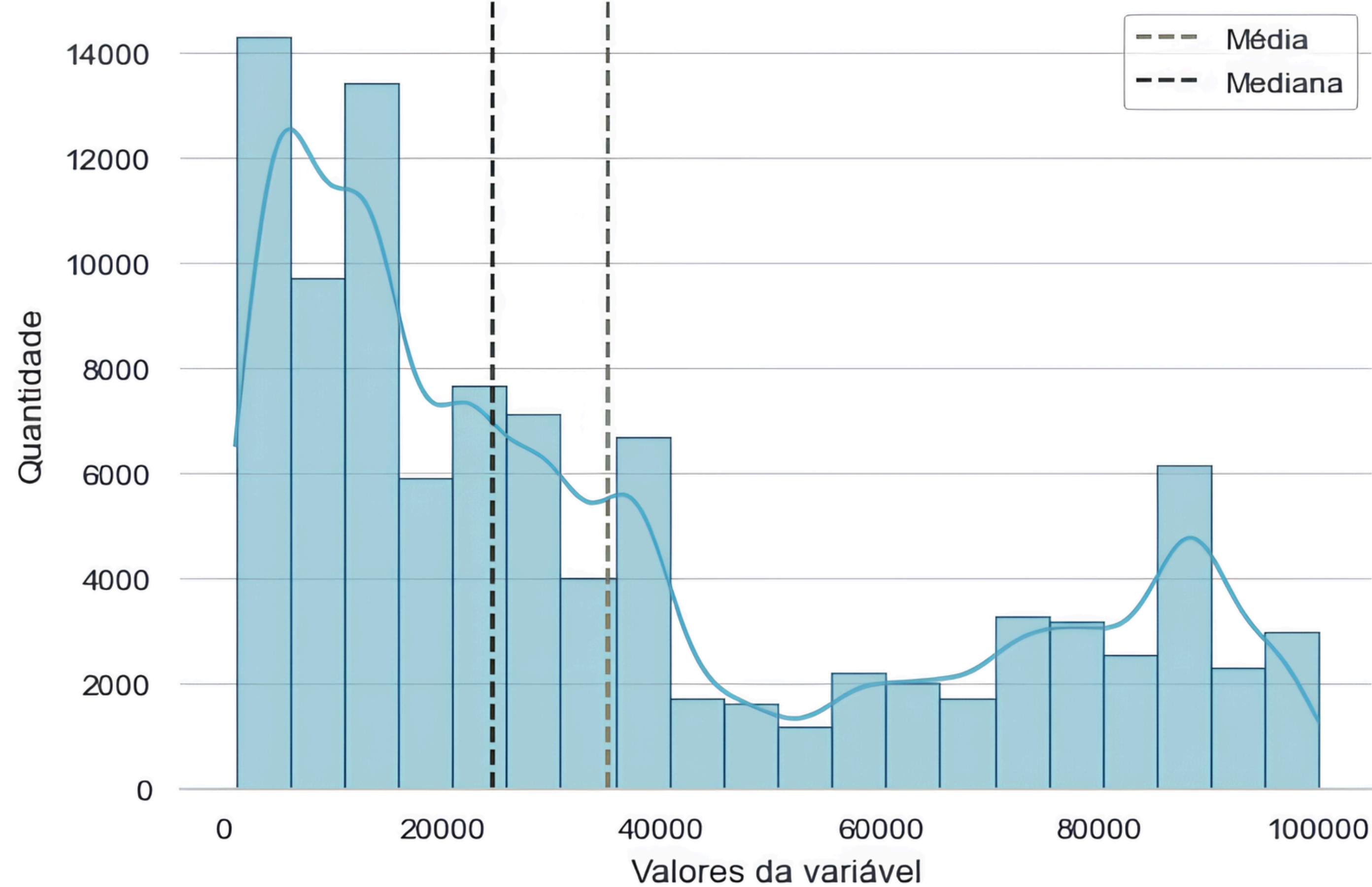


VISÃO GERAL

Este trabalho se propõe a realizar uma análise estatística detalhada sobre o conjunto de dados de vendas do comércio eletrônico brasileiro disponível no Kaggle. O foco está em entender as características gerais dos dados e responder questões fundamentais que podem orientar tanto decisões técnicas quanto de negócios.

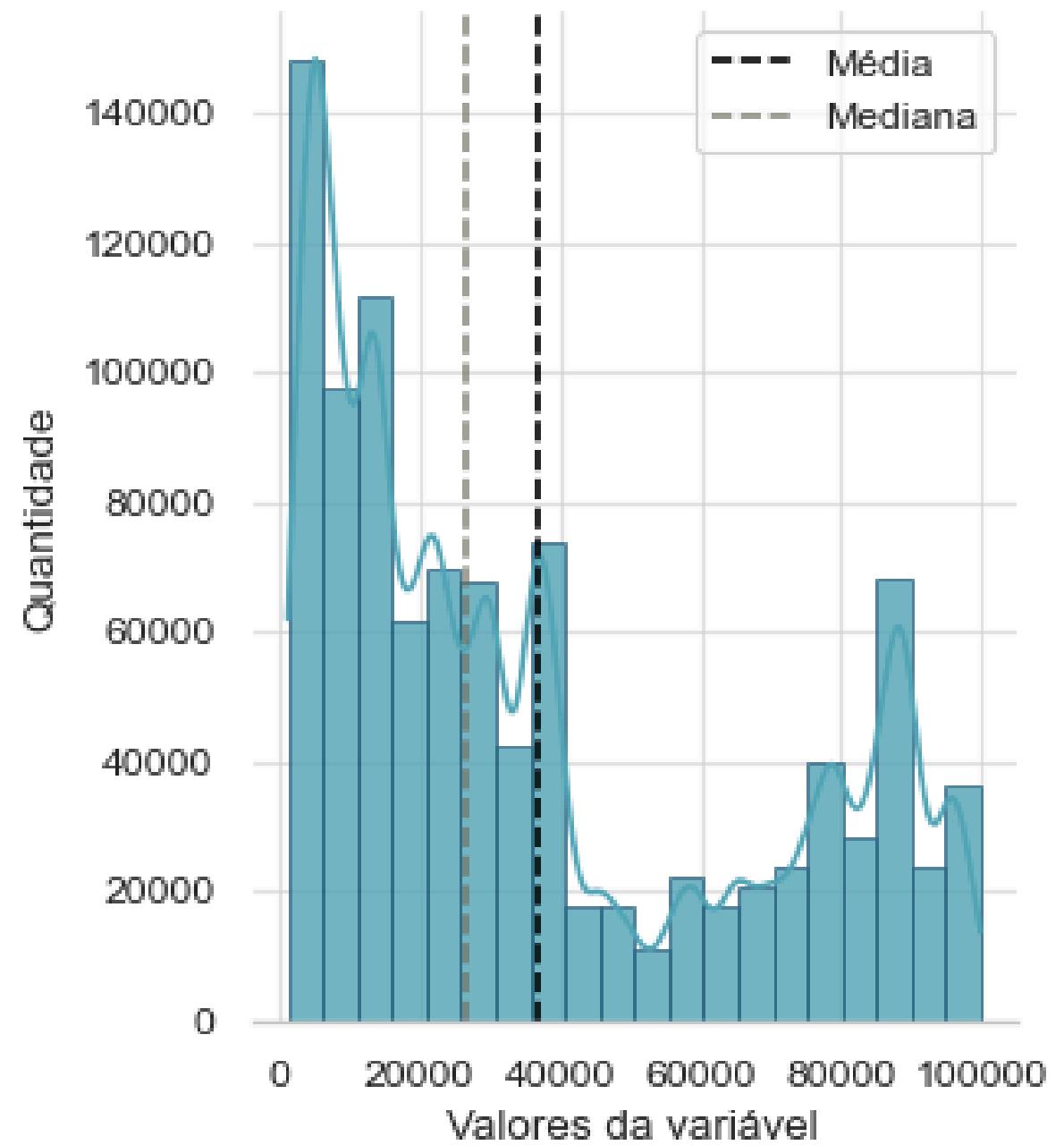
Alguns desafios foram os dados nulos encontrados nas variáveis: `review_comment_title, review_comment_message, order_approve_d_at, order_delivered_carrier_date, order_delivered_customer_date, product_category_name, product_name_length, product_description_length, product_photos_qty, product_weight_g, product_length_cm, product_height_cm, product_width_cm,`

Distribuição da variável customer_zip_code_prefix

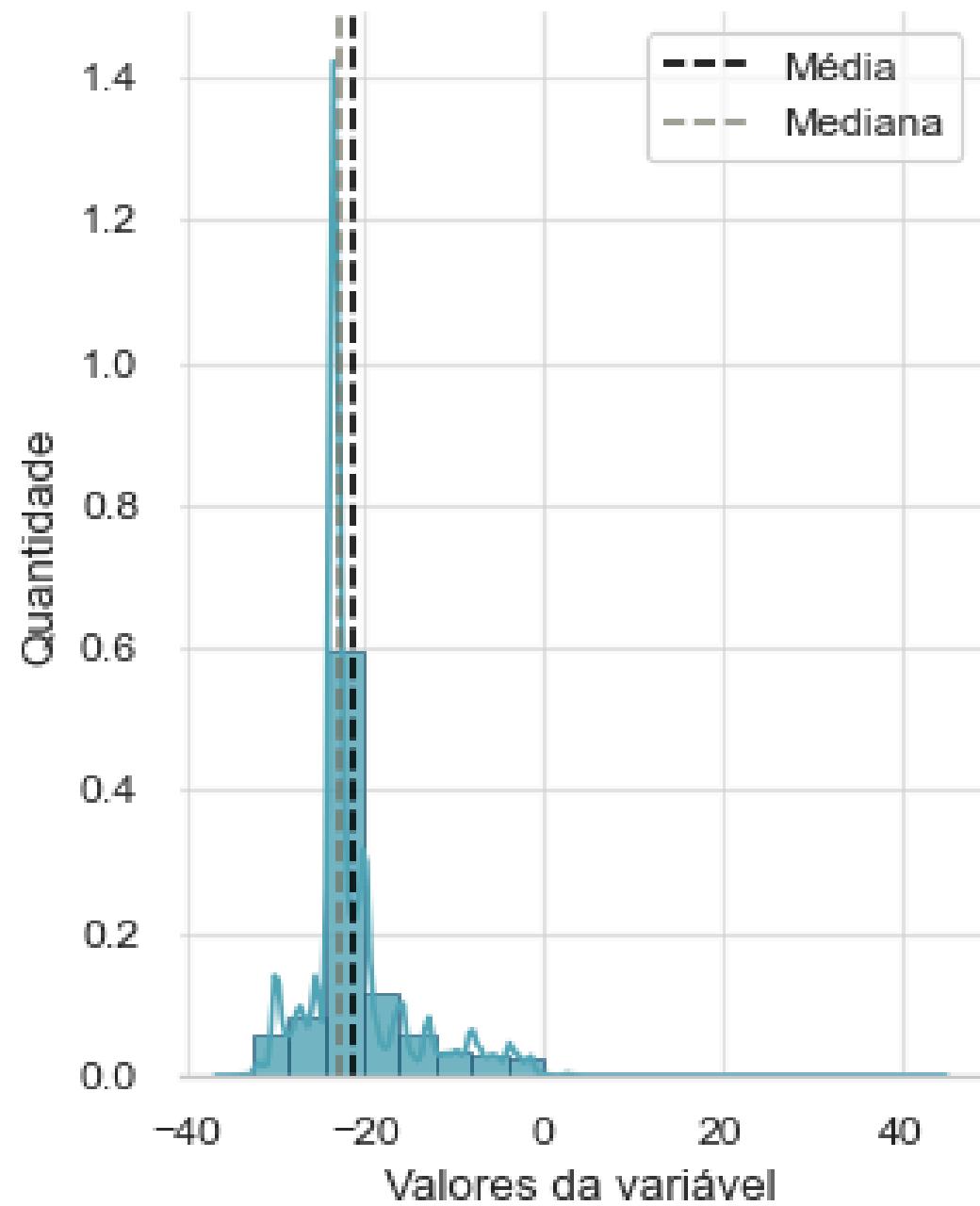


Distribuição das variáveis numéricas de df_geolocalizacao

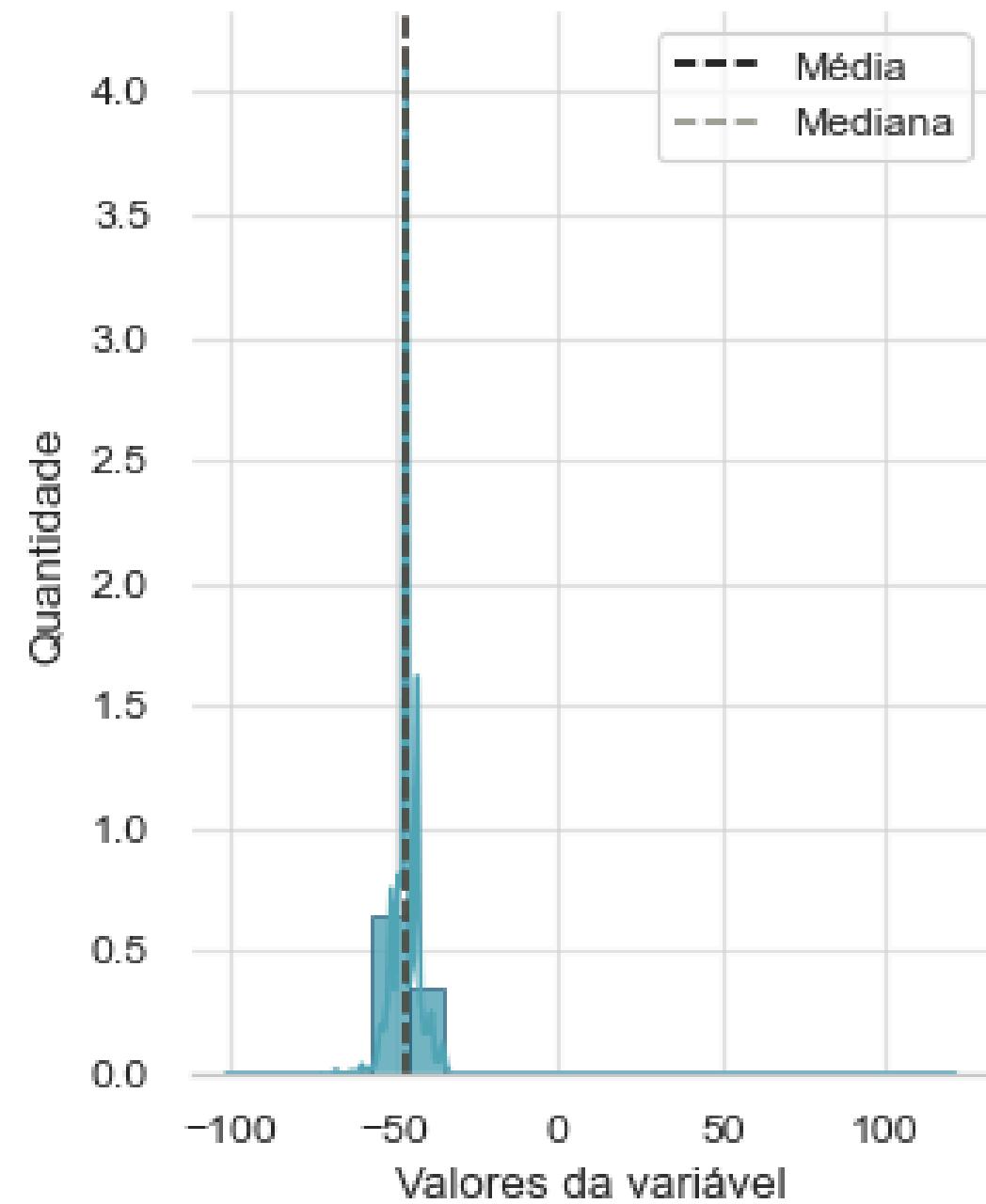
Distribuição de geolocation_zip_code_prefix



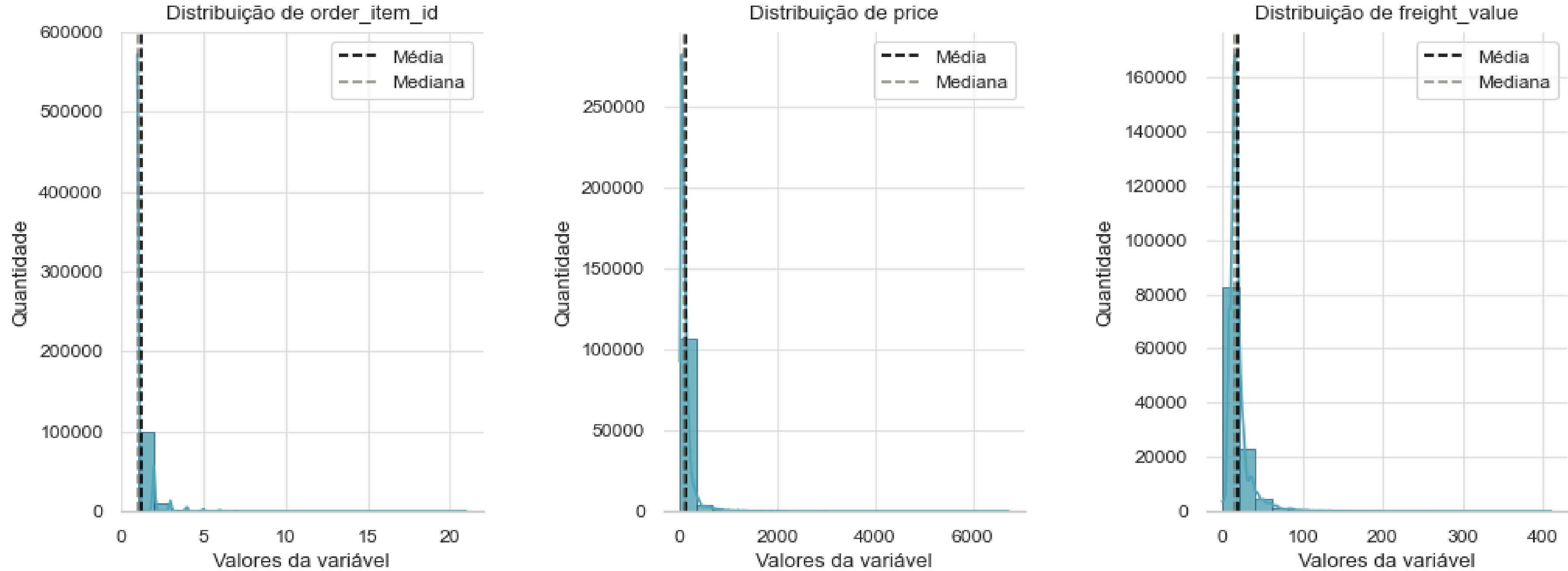
1e6 Distribuição de geolocation_lat



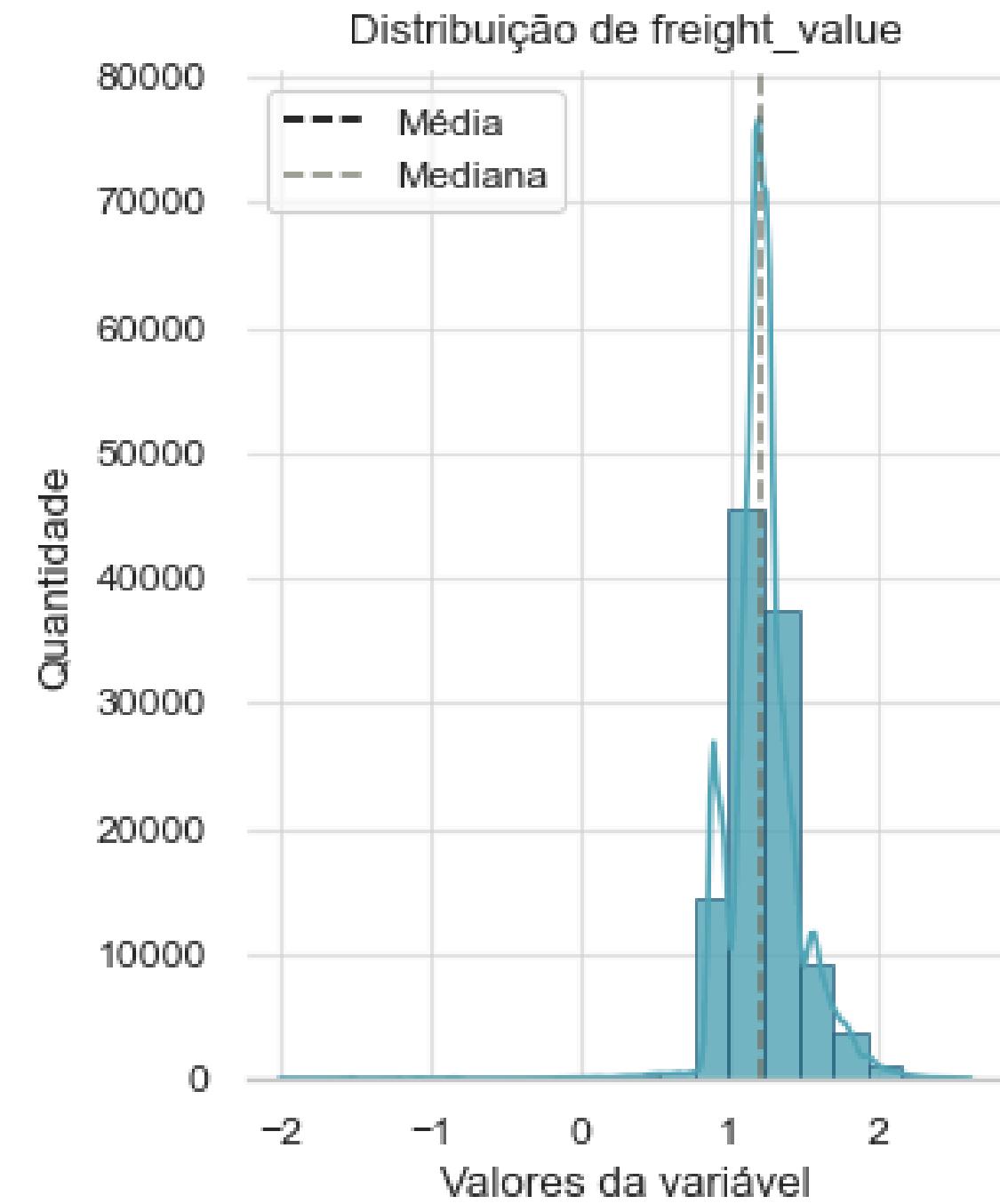
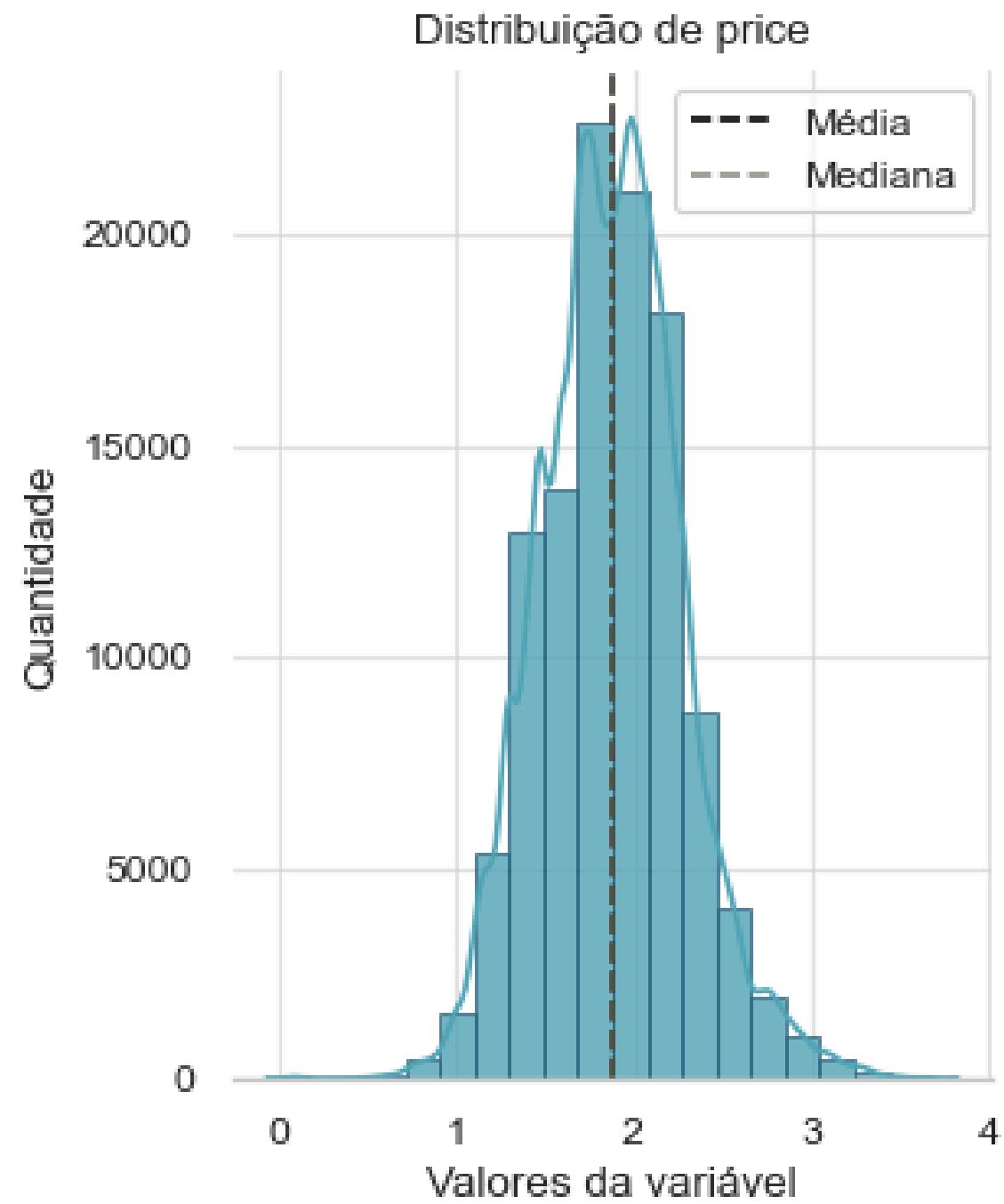
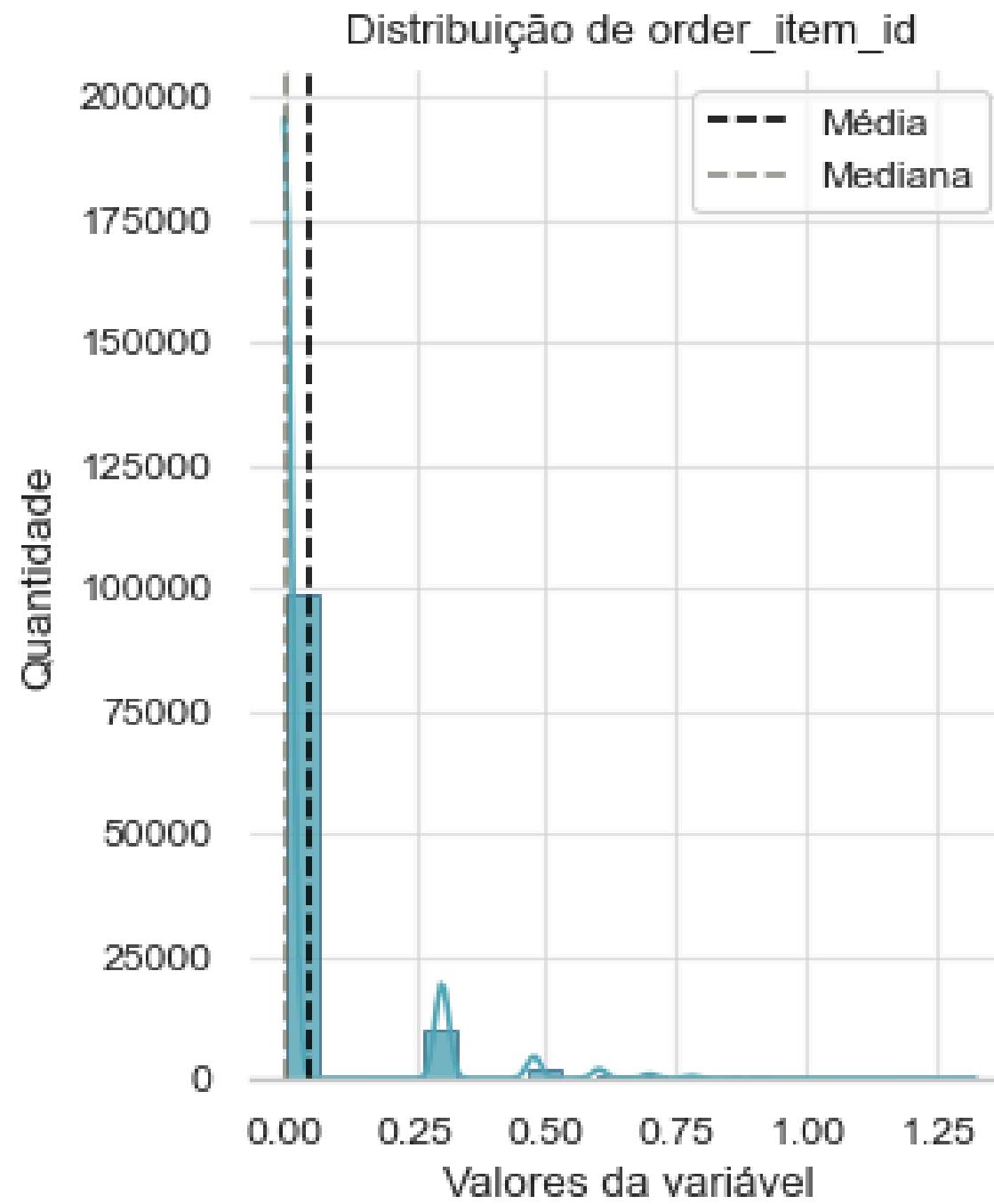
1e6 Distribuição de geolocation_lng



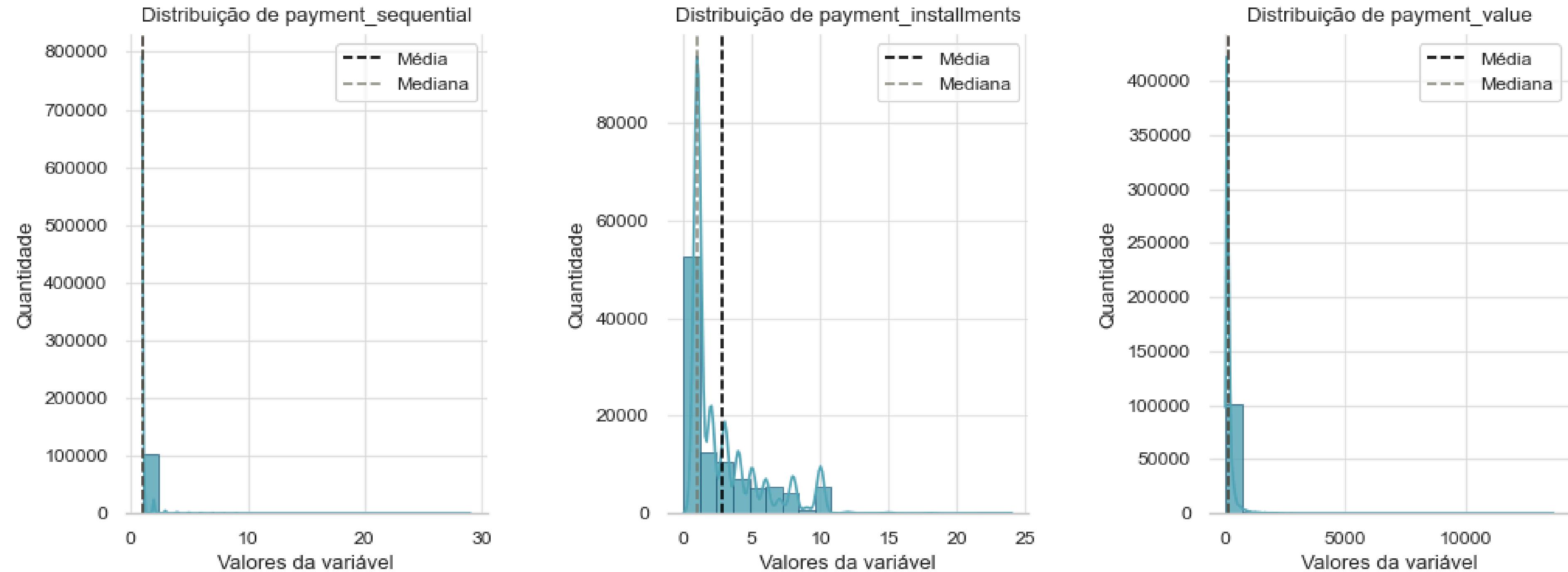
Distribuição das variáveis numéricas de df_pedidos_itens



Distribuição das variáveis numéricas (logaritimizadas na base 10) de df_pedidos_itens

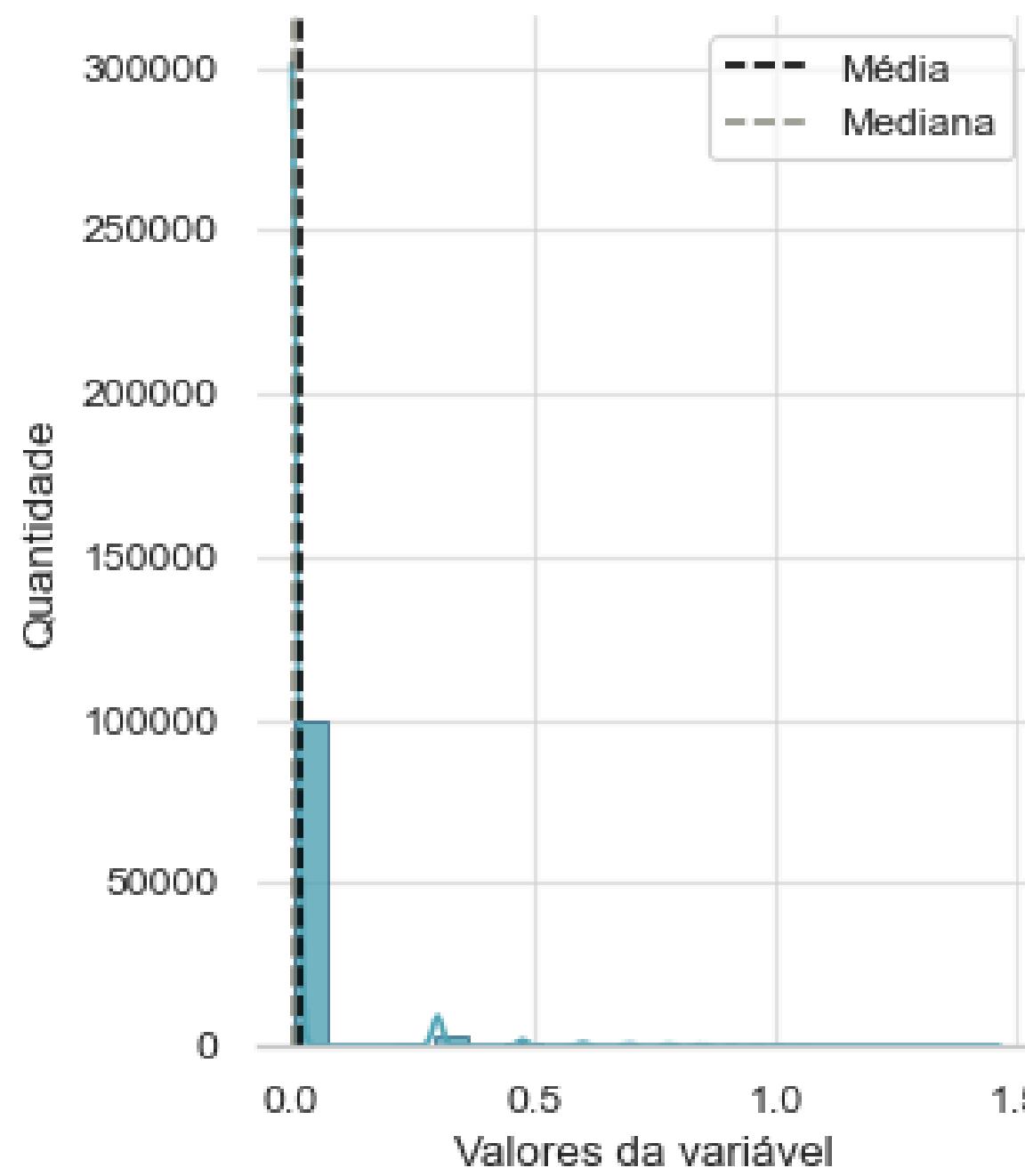


Distribuição das variáveis numéricas de df_pagamentos

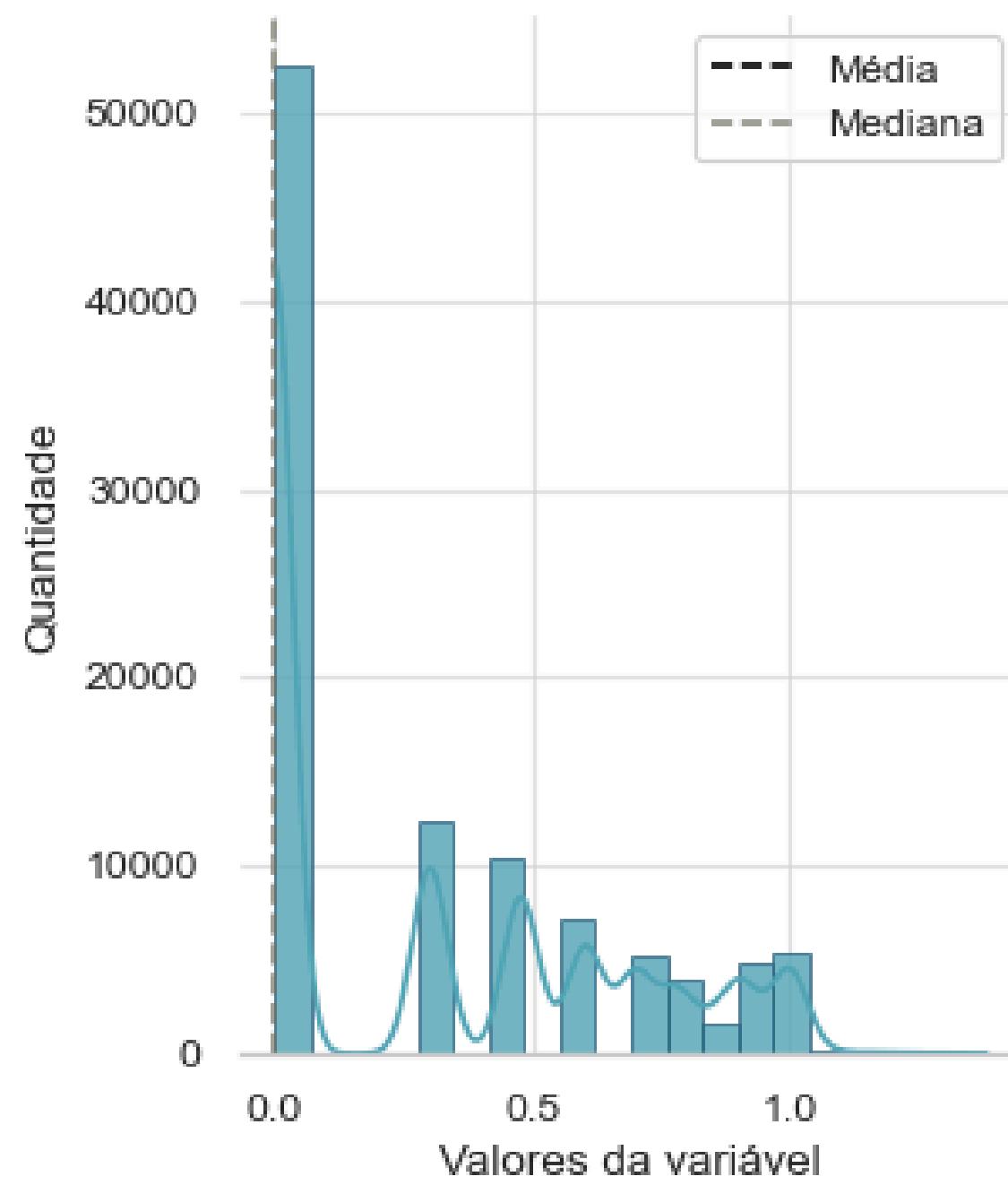


Distribuição das variáveis numéricas (logaritimizadas na base 10) de df_pedidos_pagamentos

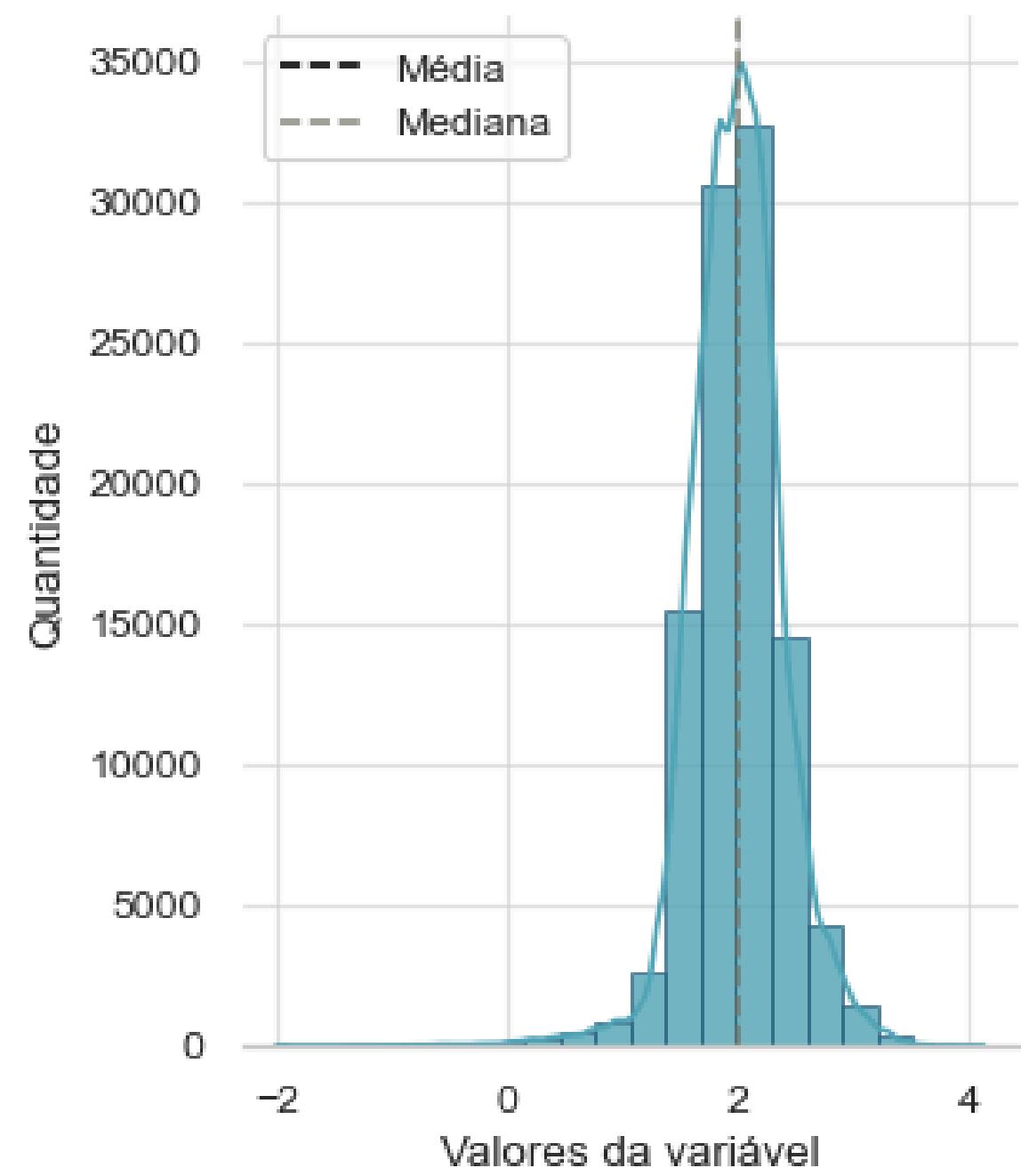
Distribuição de payment_sequential



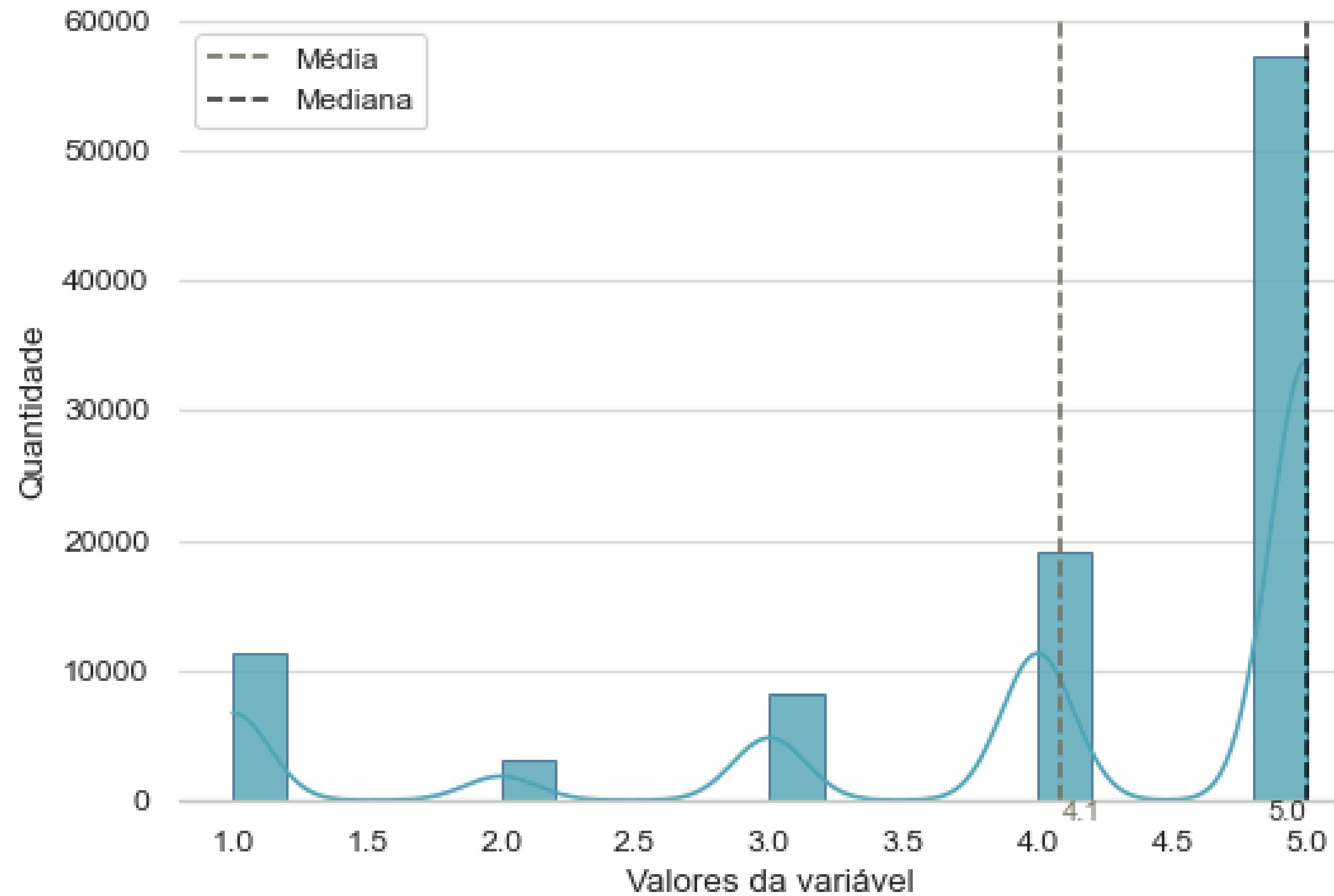
Distribuição de payment_installments



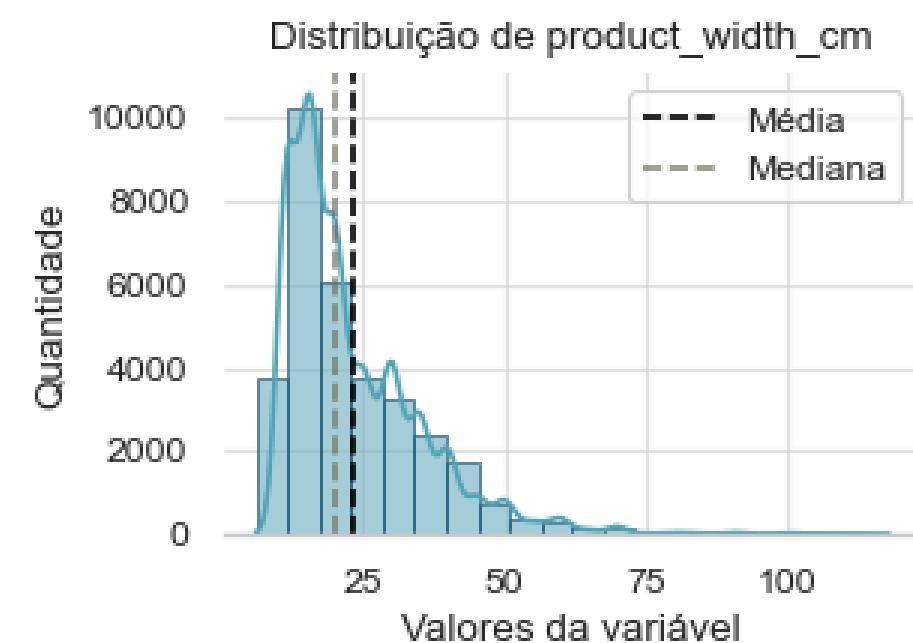
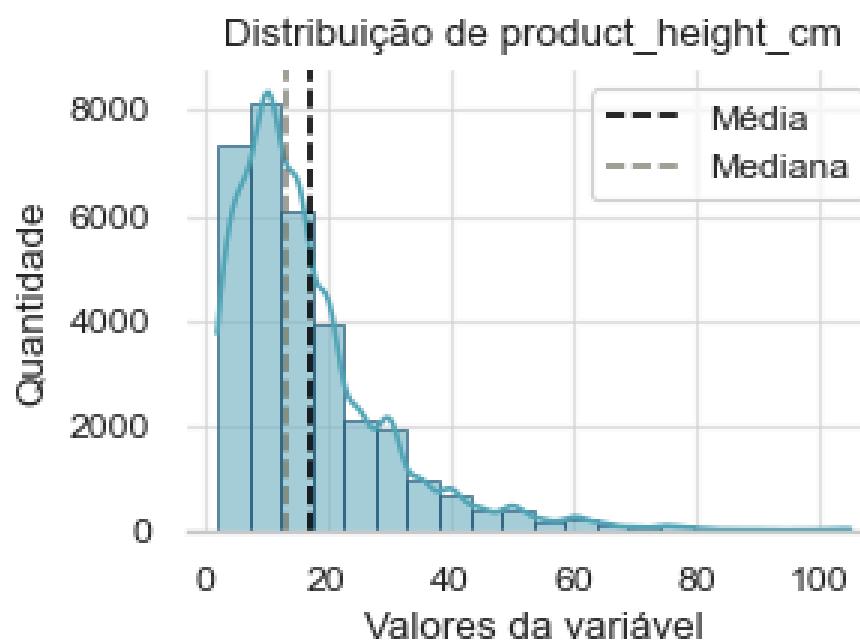
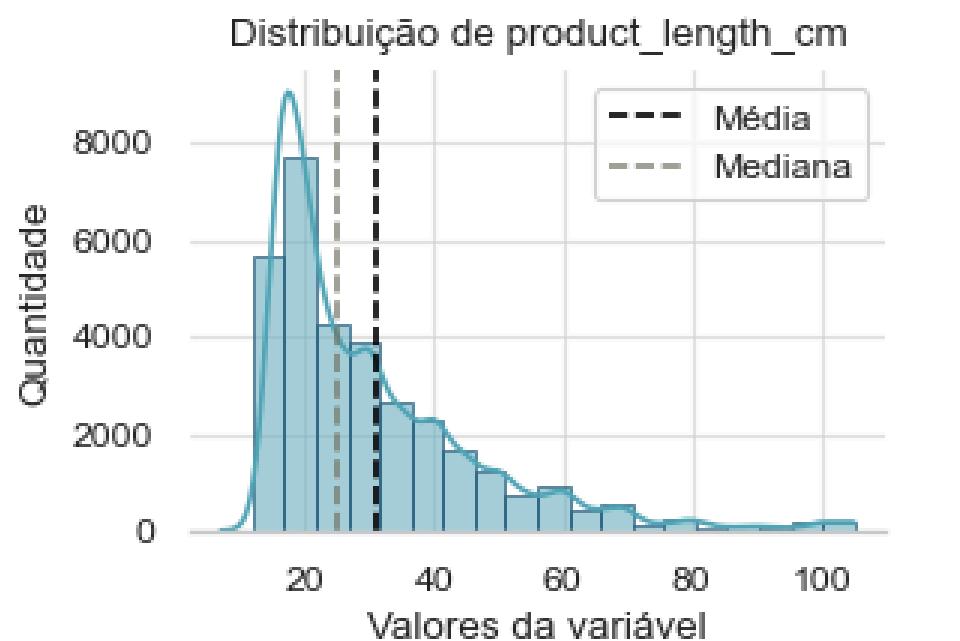
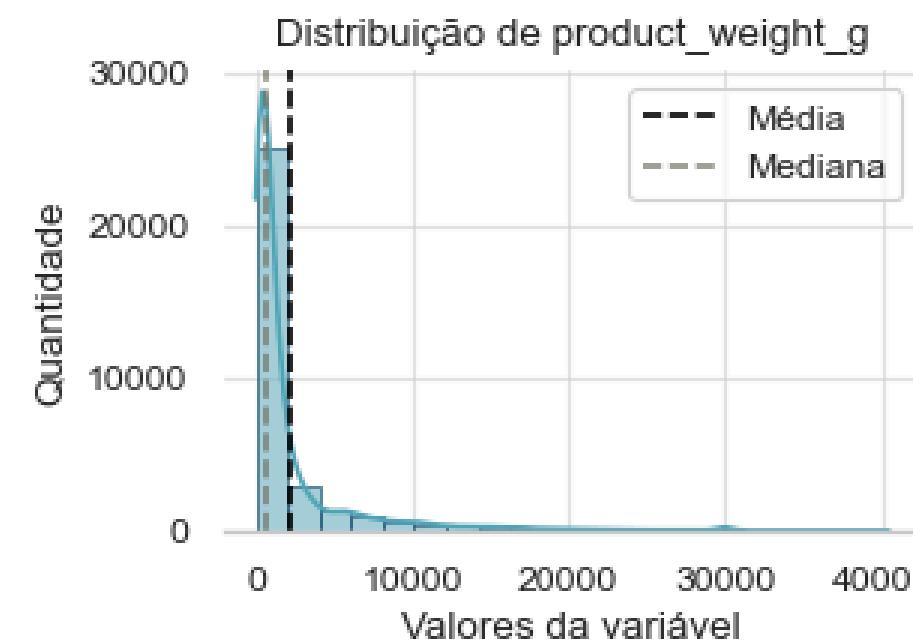
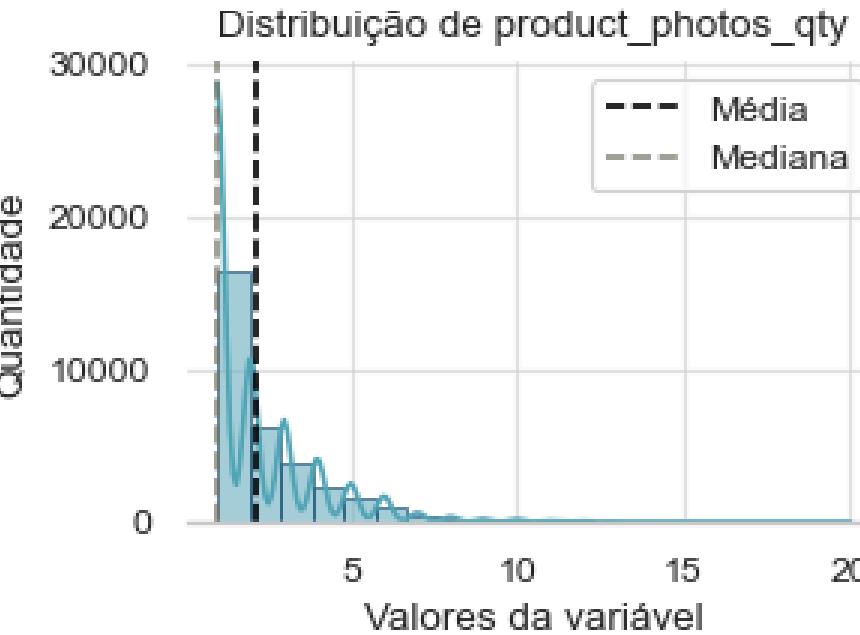
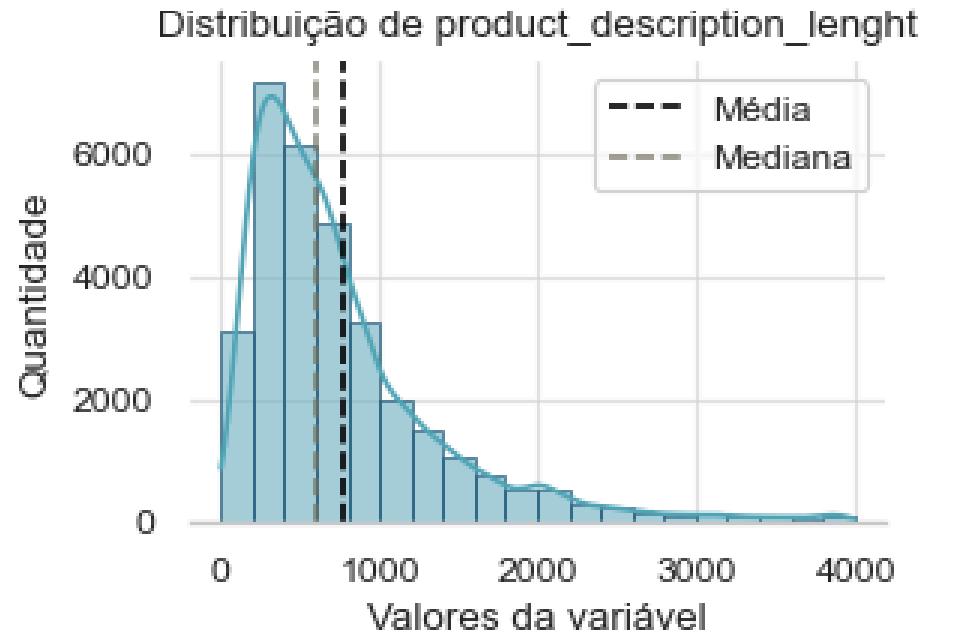
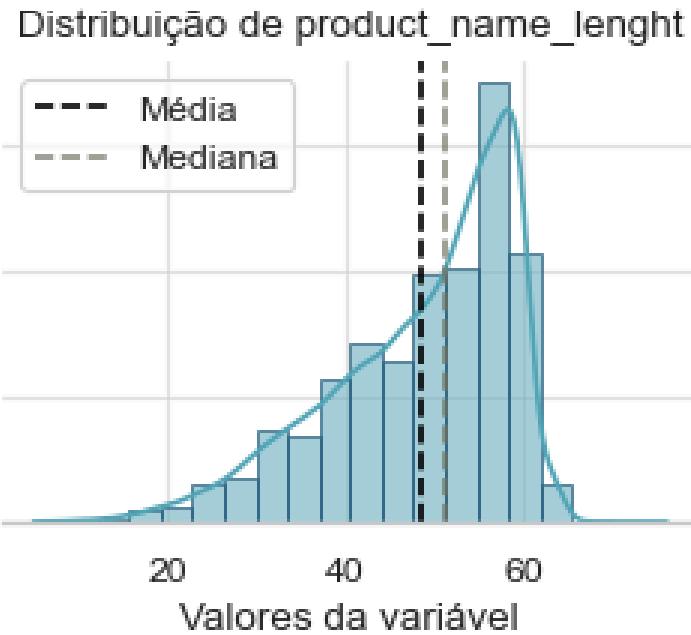
Distribuição de payment_value



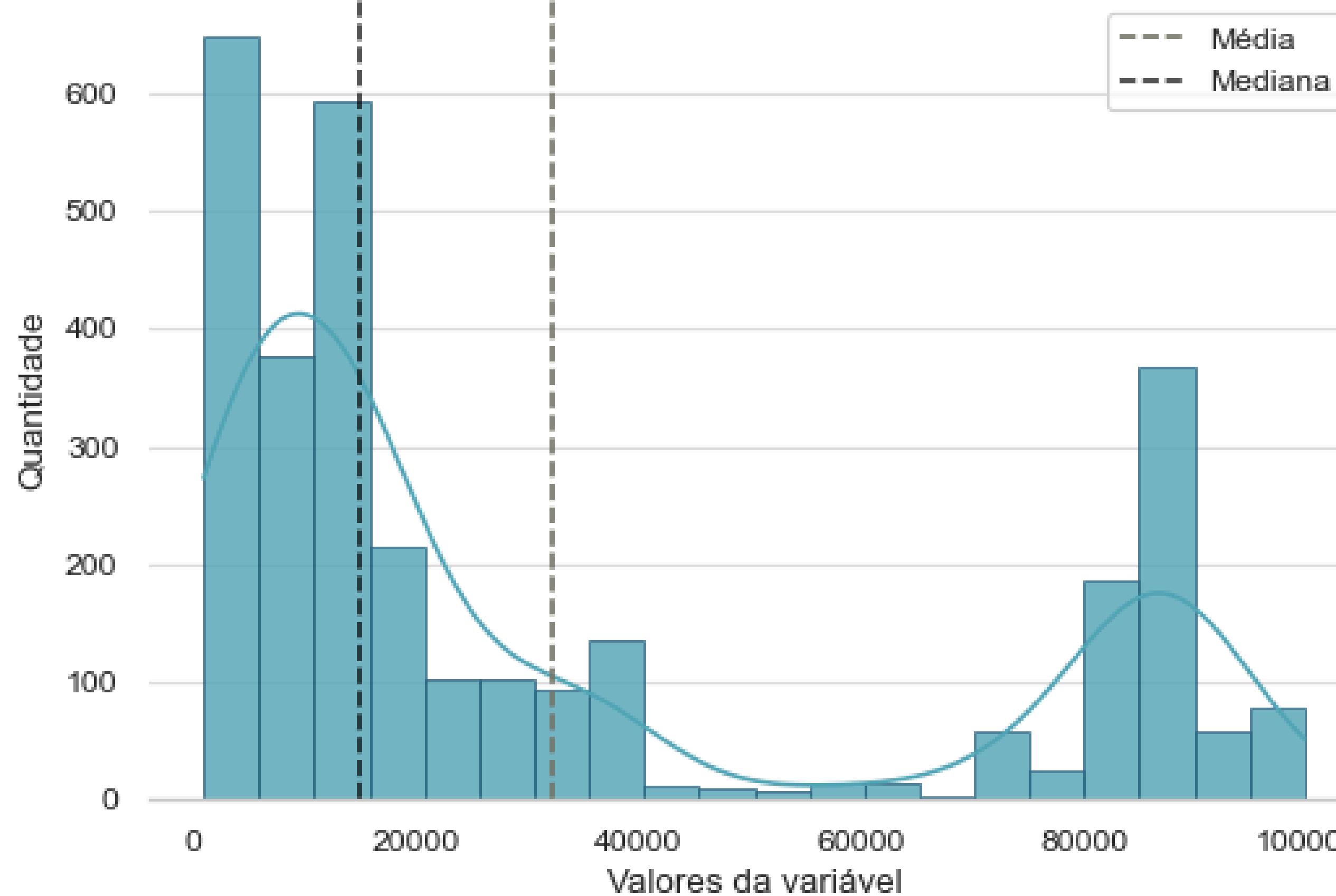
Distribuição da variável review_score



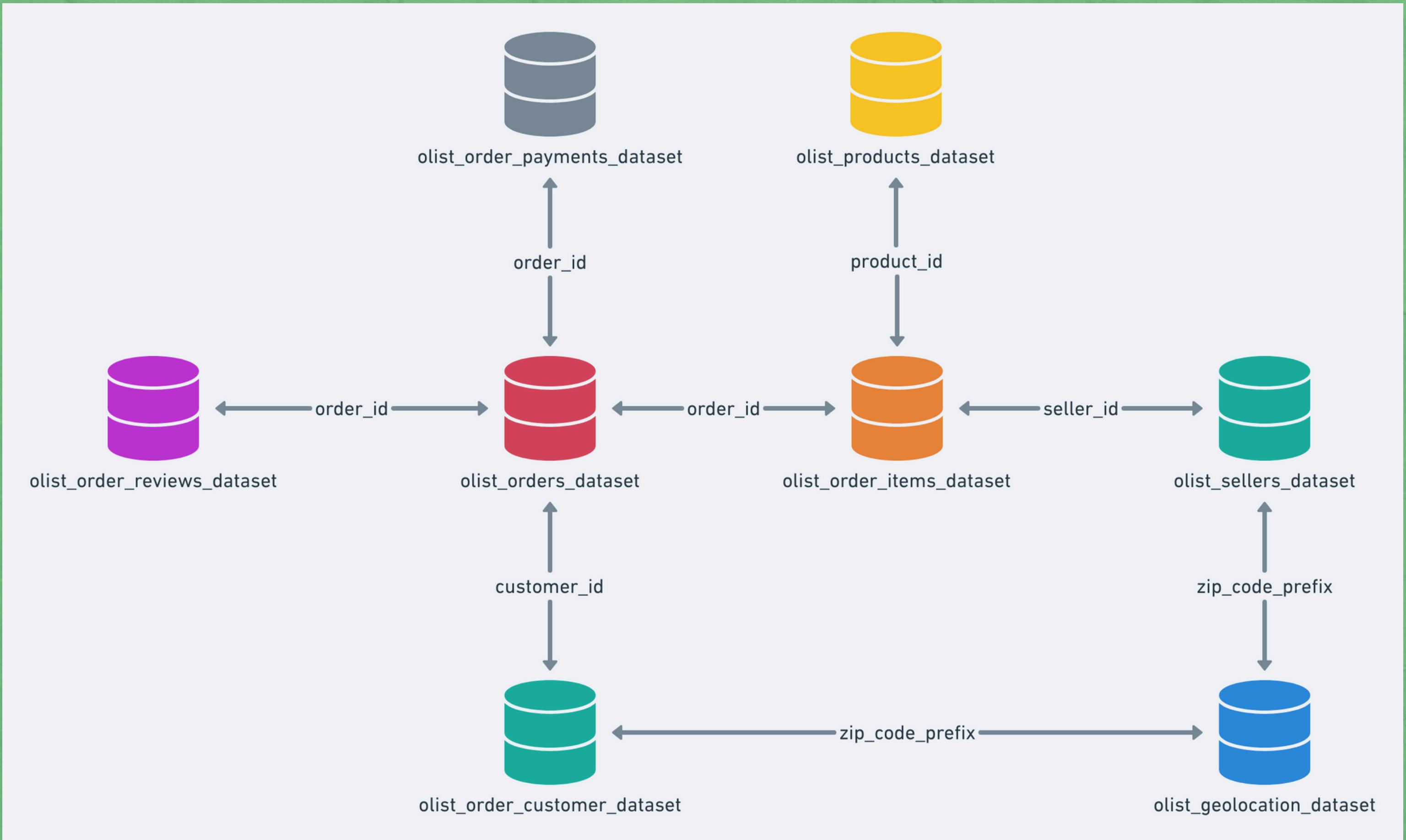
Distribuição das variáveis numéricas de df_produtos



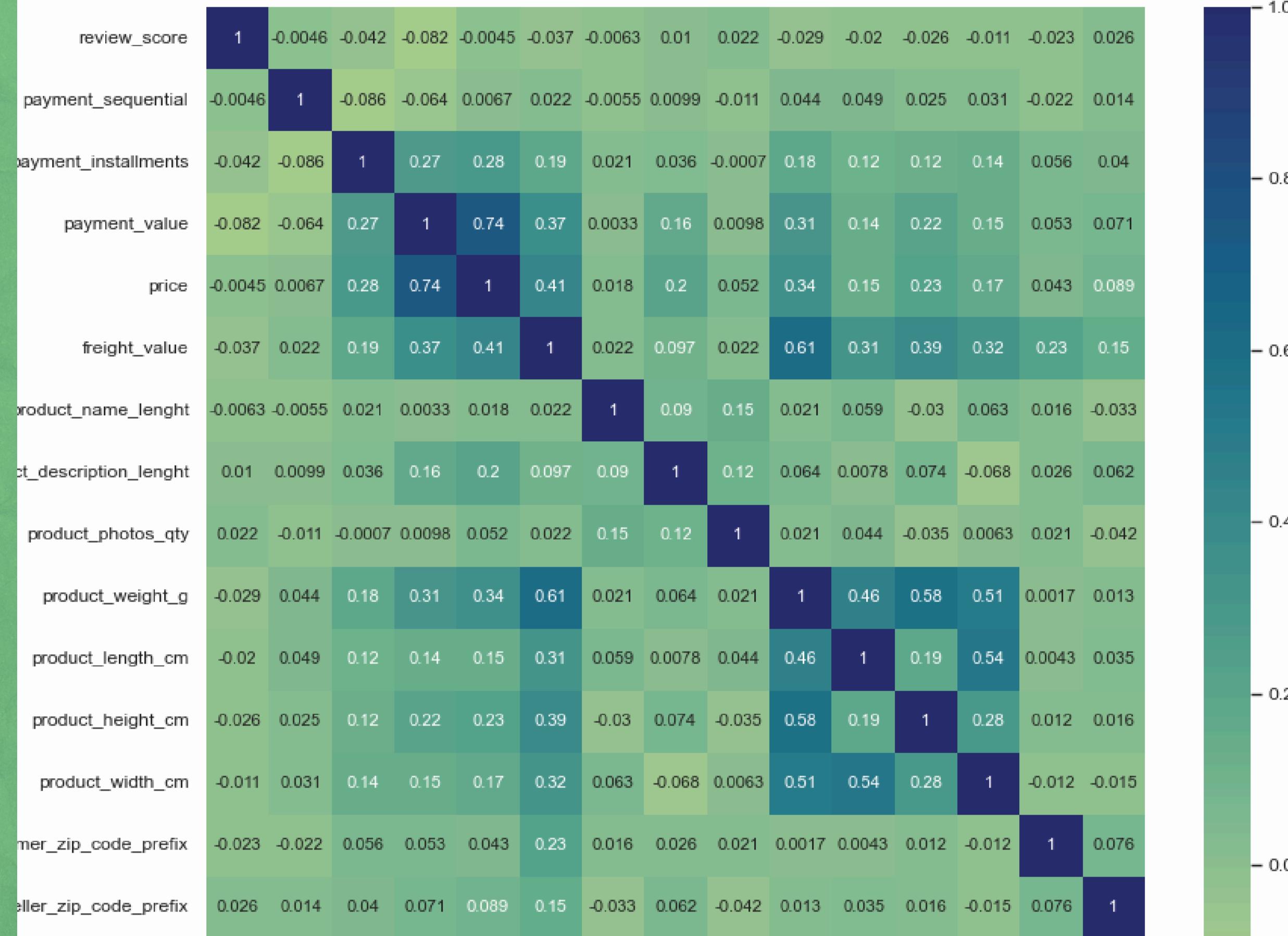
Distribuição da variável seller_zip_code_prefix



Combinando os datasets para análise de correlação



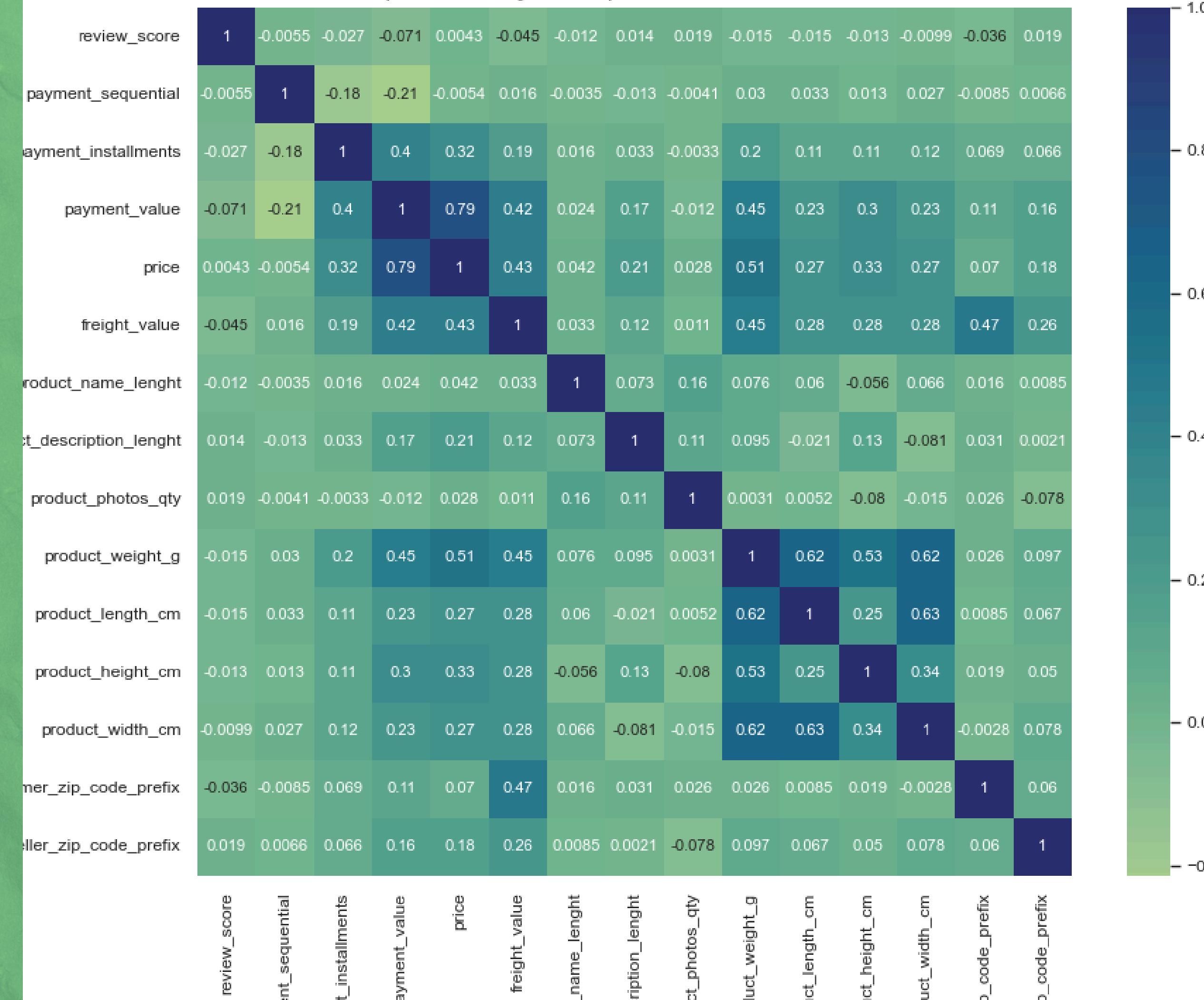
Heatmap da Correlação de Pearson das Variáveis Numéricas



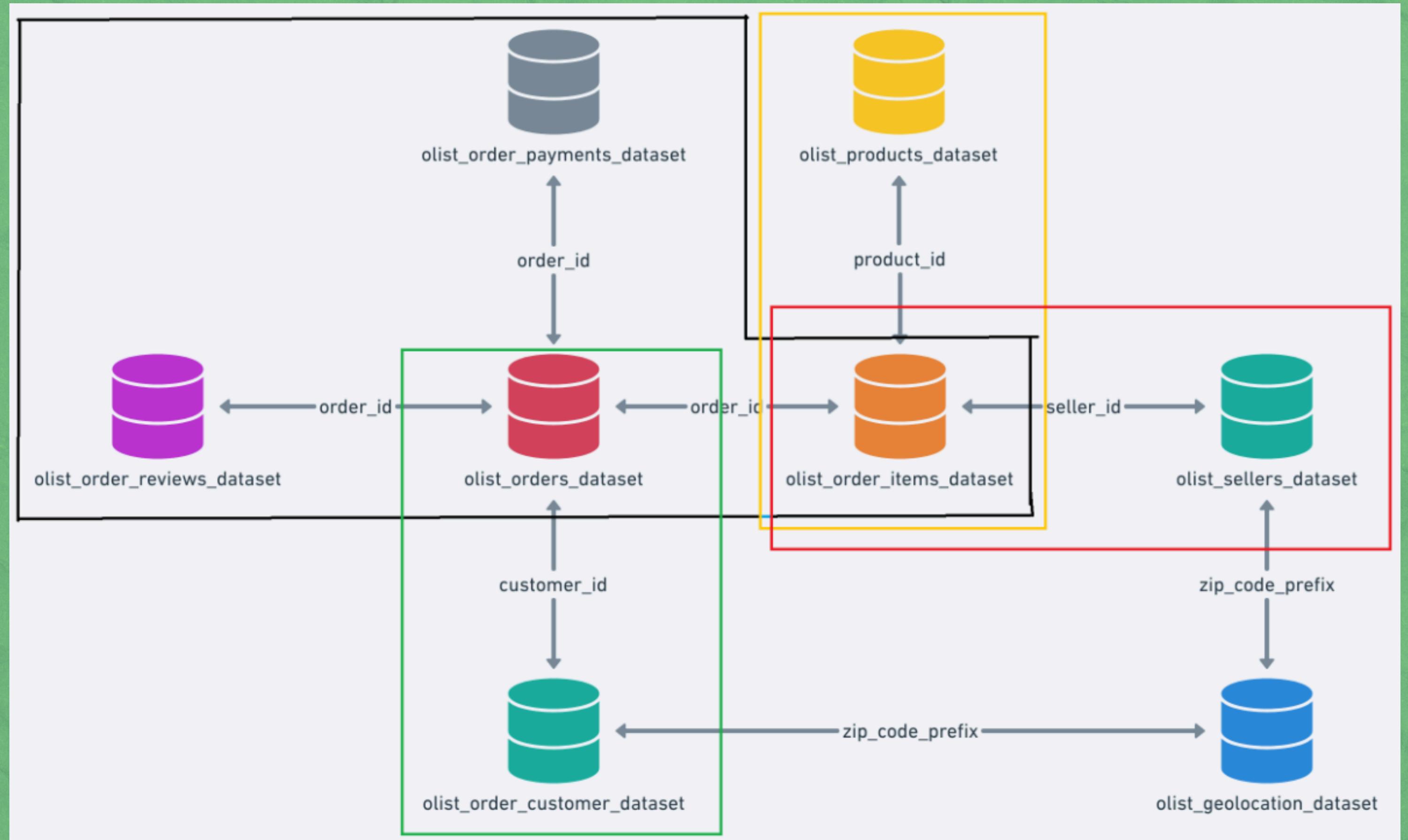
review_score payment_sequential payment_installments payment_value price freight_value product_name_length product_description_length product_photos_qty product_weight_g product_length_cm product_height_cm product_width_cm buyer_zip_code_prefix seller_zip_code_prefix



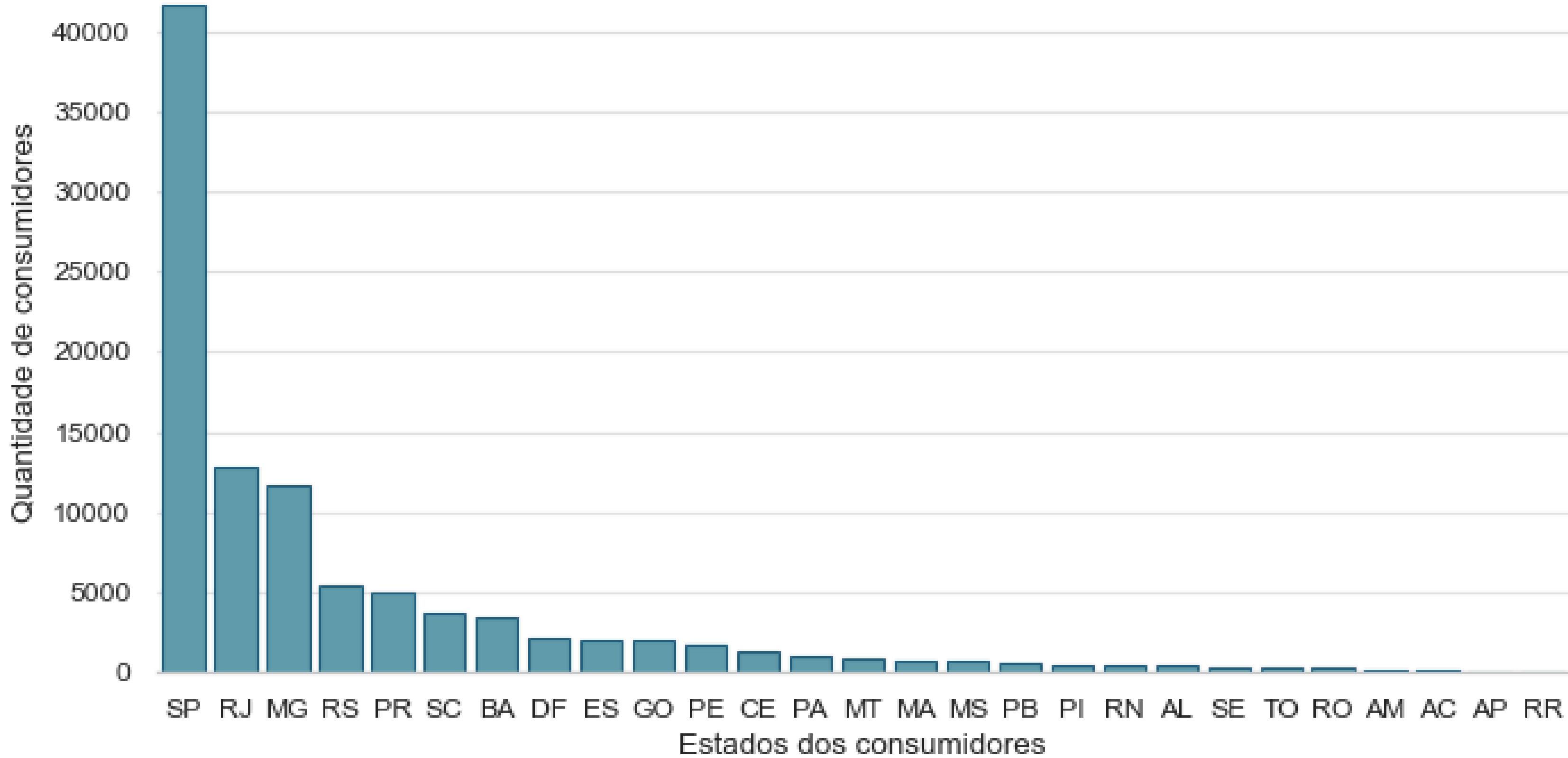
Heatmap de Correlação da Spearman das Variáveis Numéricas



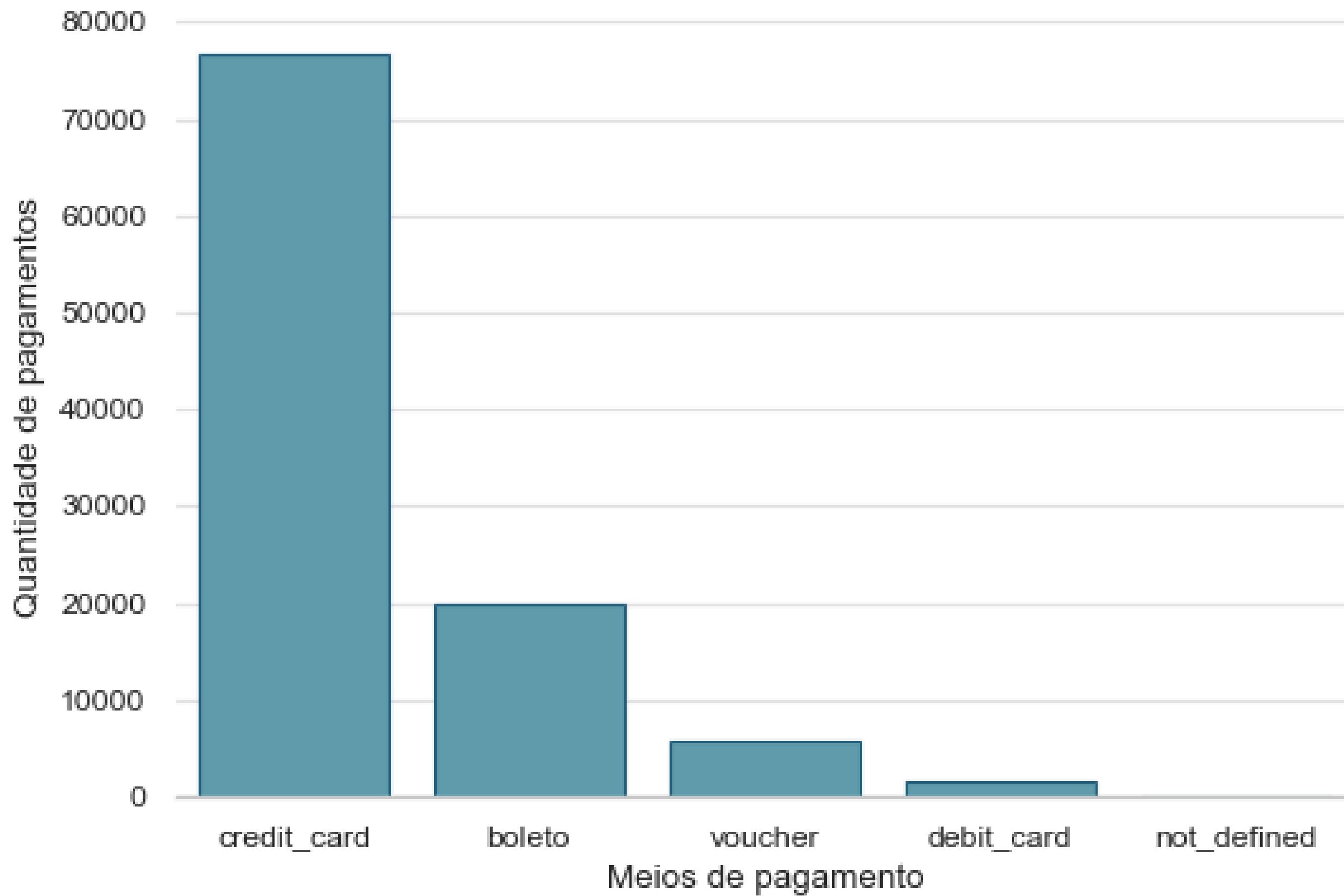
Combinando os datasets para análise de correlação



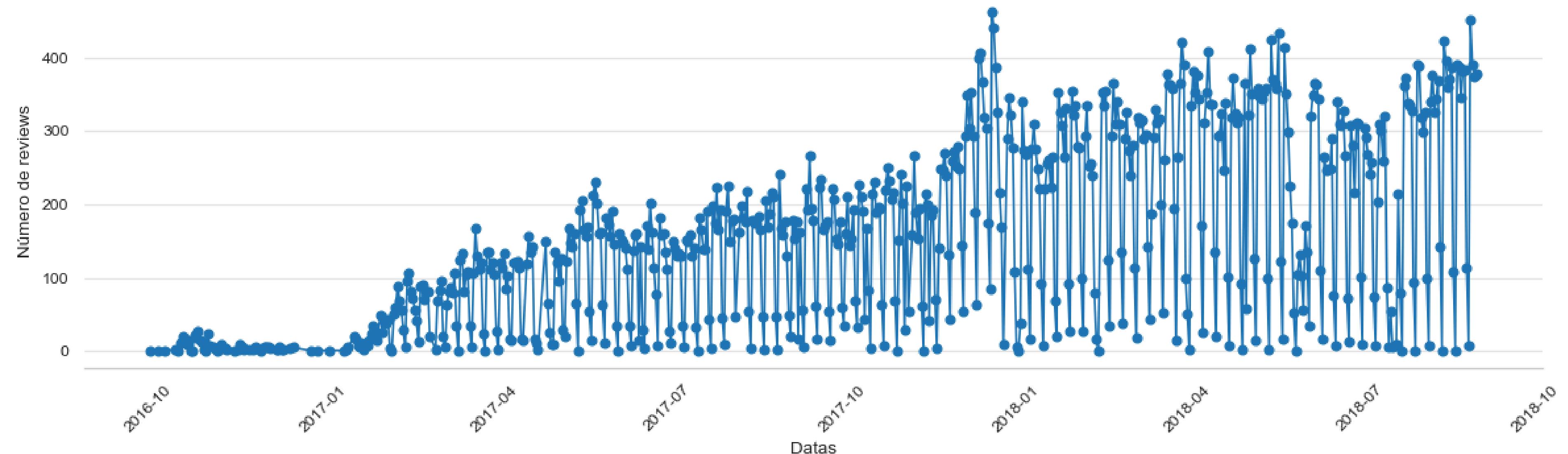
Distribuição das UFs dos Consumidores



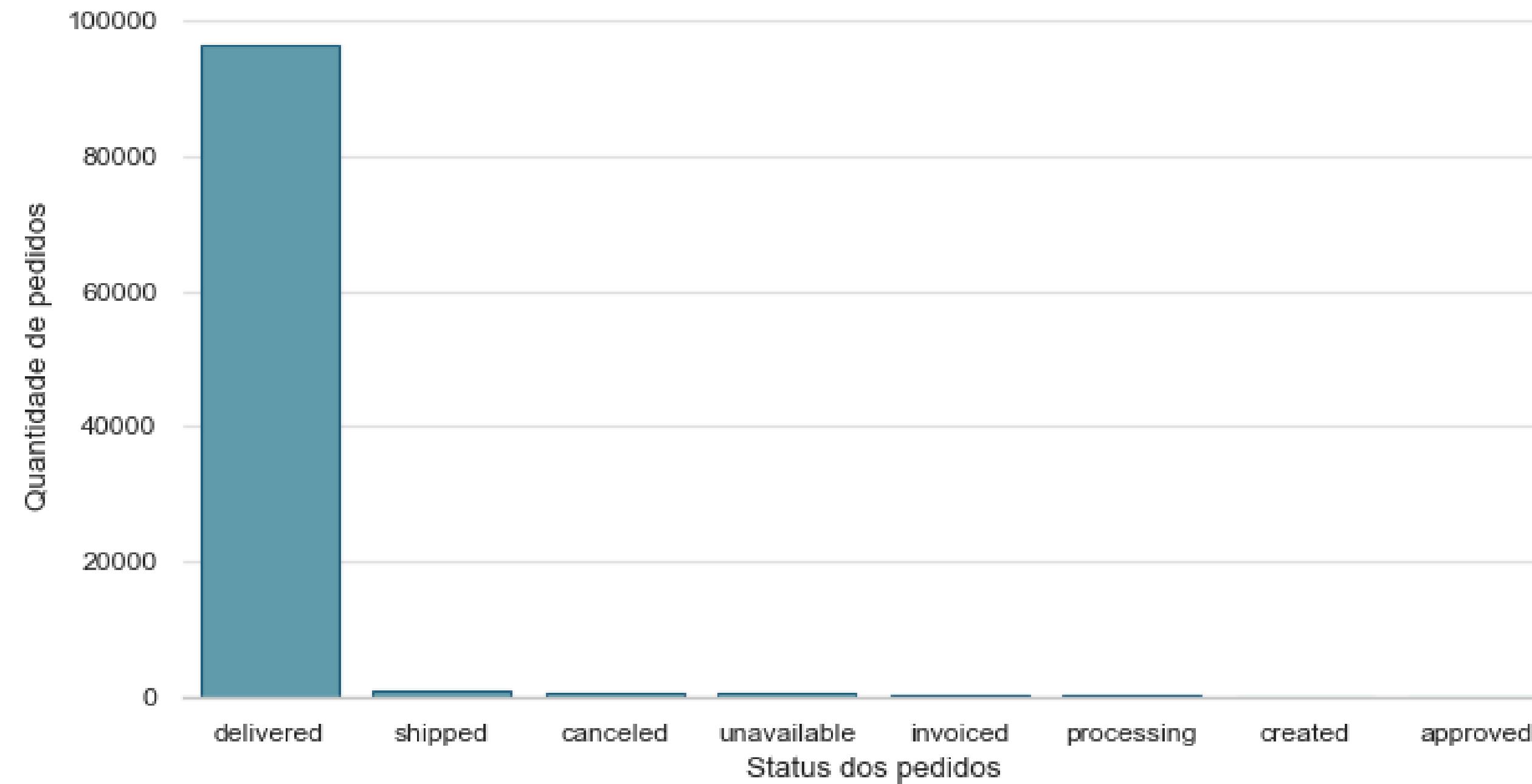
Distribuição do número de pagamentos por meios de pagamento



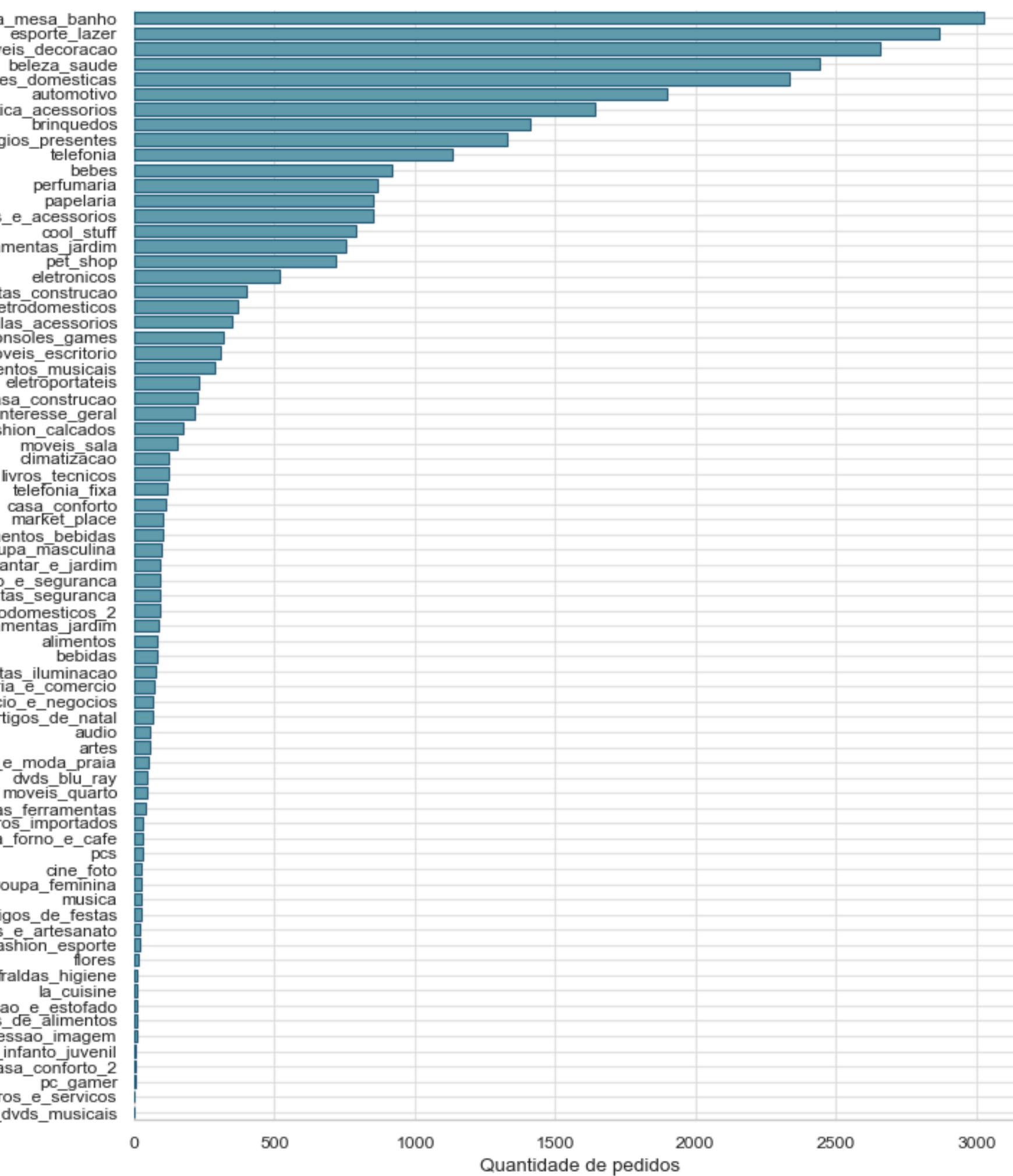
Número de reviews por data de criação



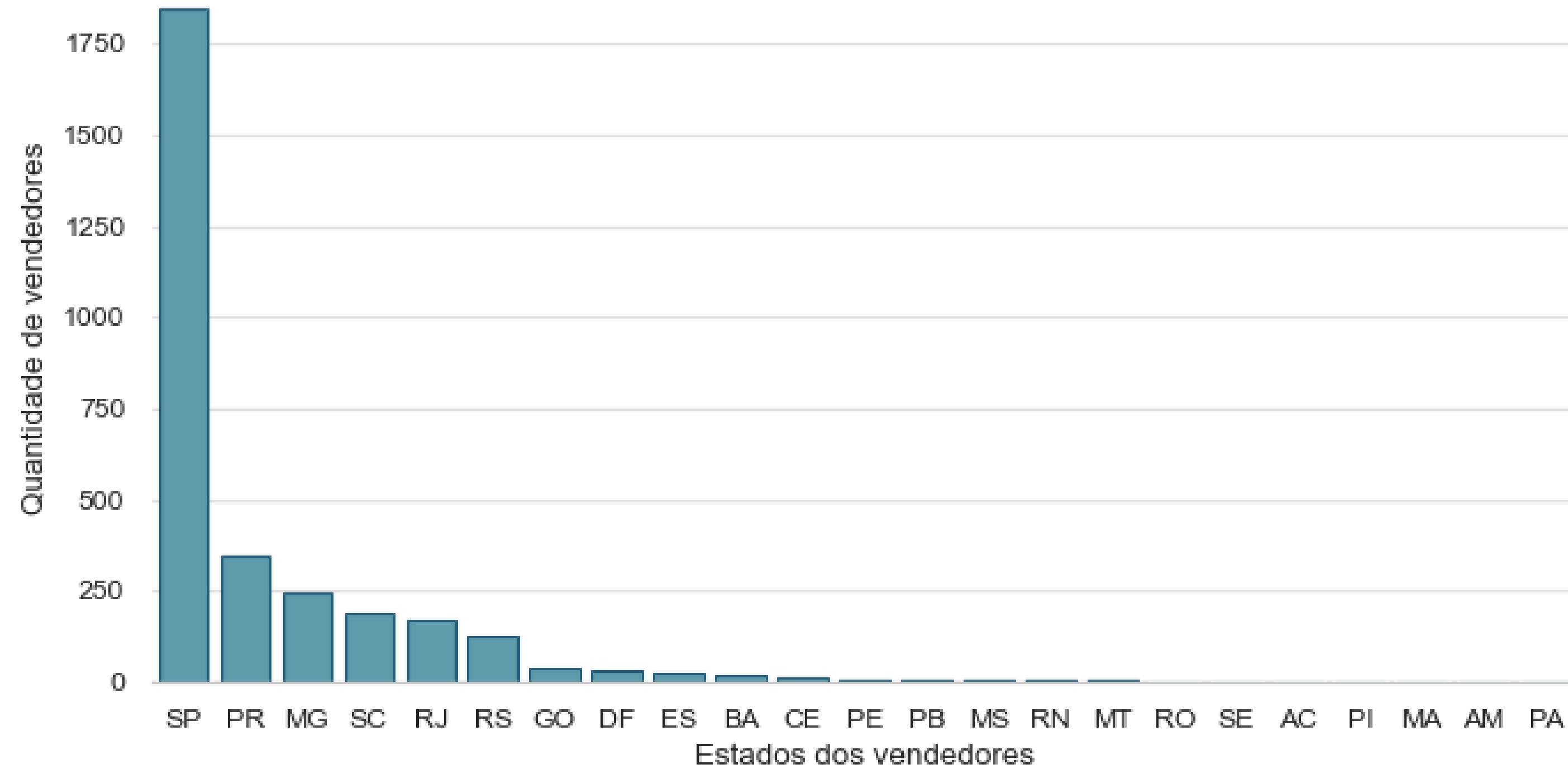
Distribuição da quantidade de pedidos por Status



Quantidades de pedidos por categoria de produto



Distribuição das UFs dos Vendedores

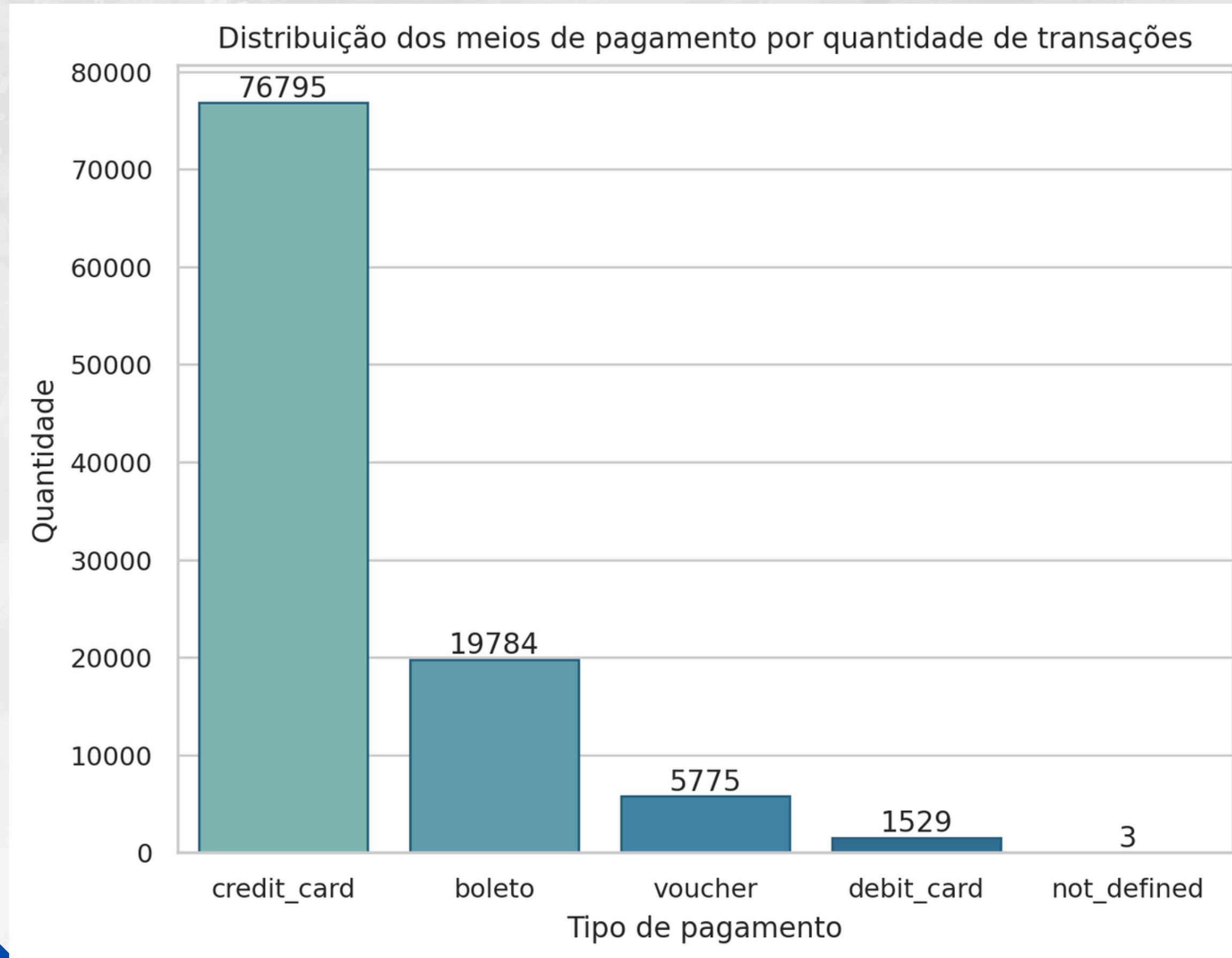


STORYTELLING

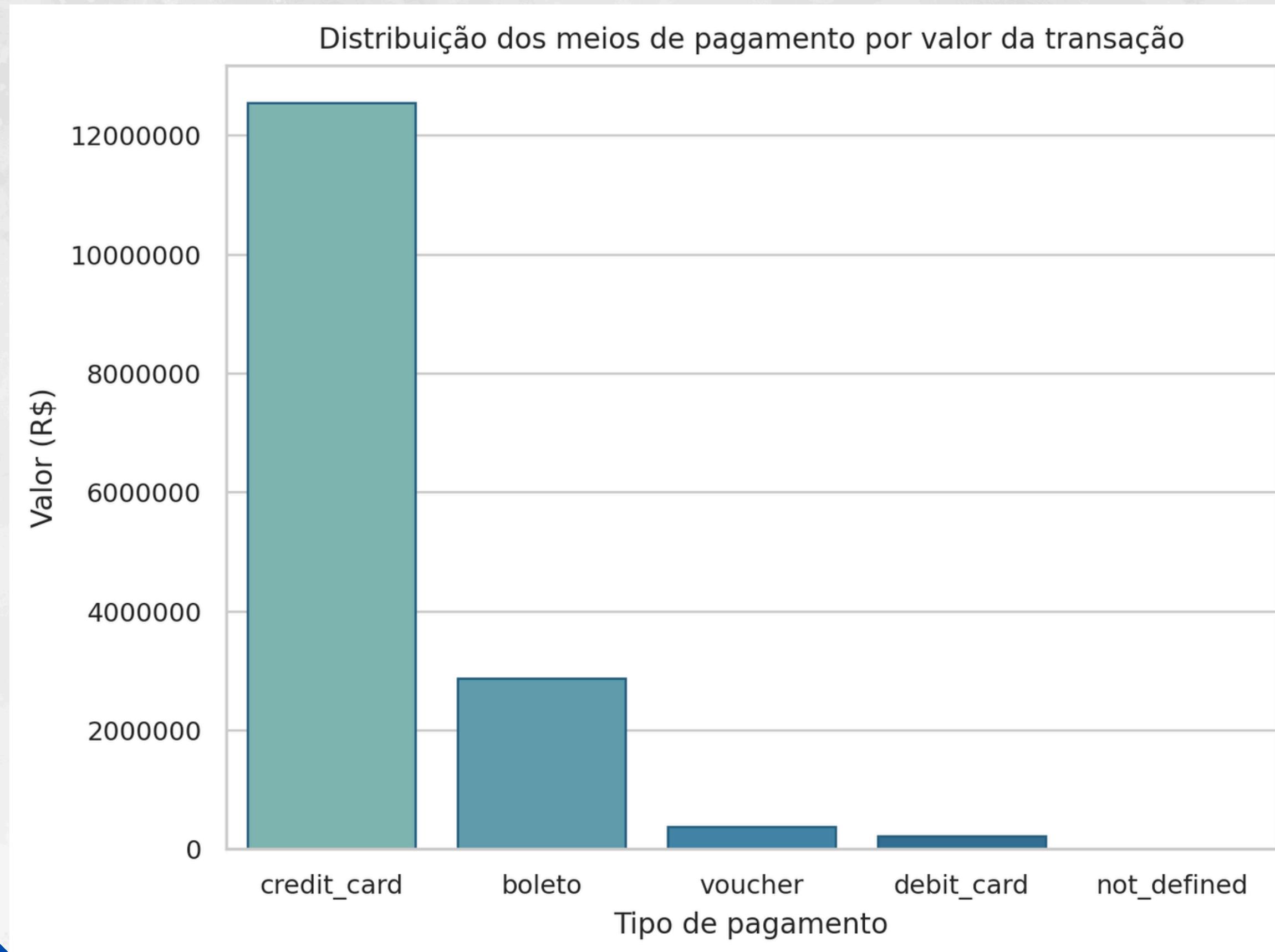
Tendências nos e-commerces brasileiros - uma análise sobre seus pedidos e os principais meios de pagamento



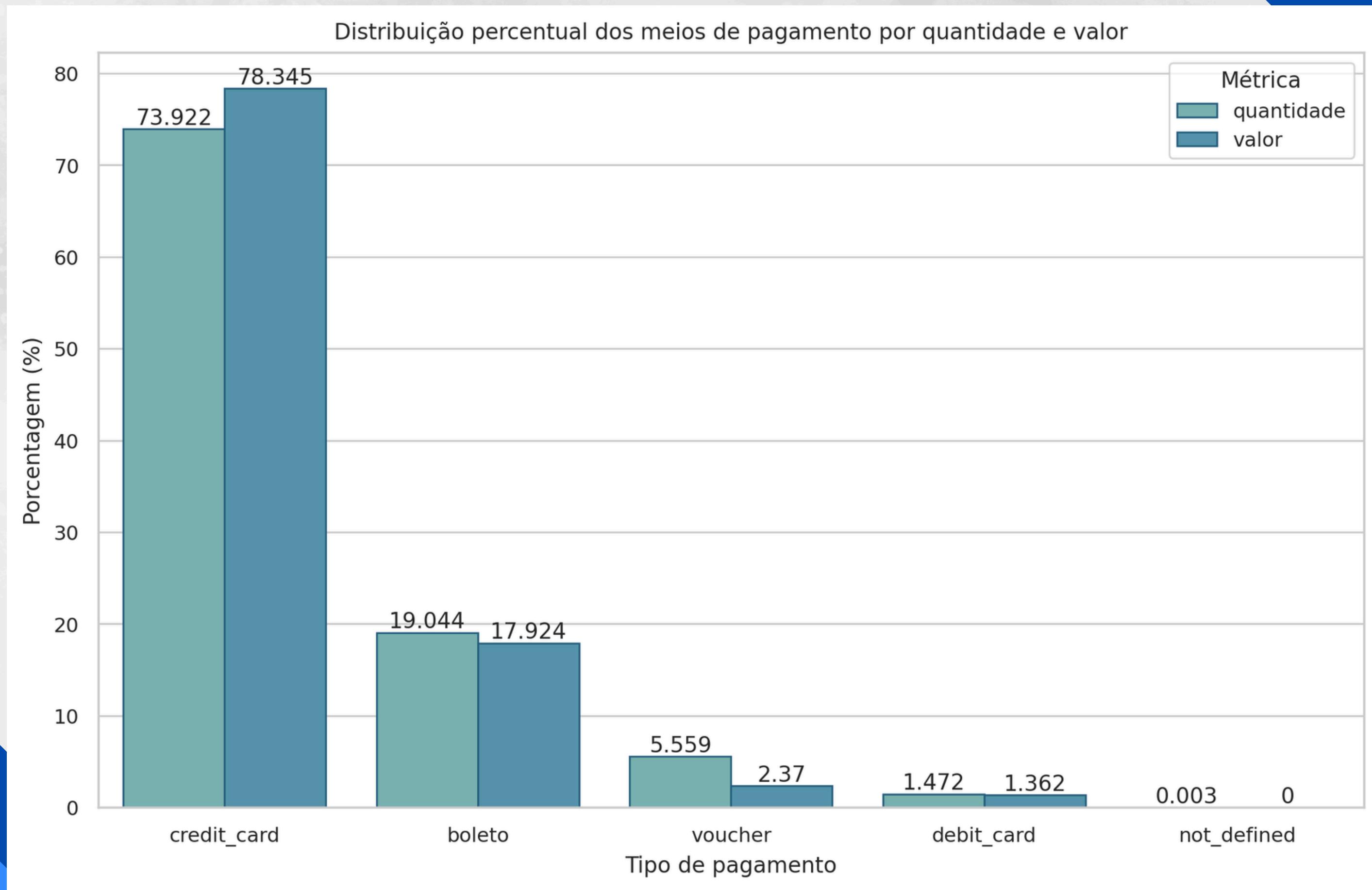
Análise sobre os meios de pagamento



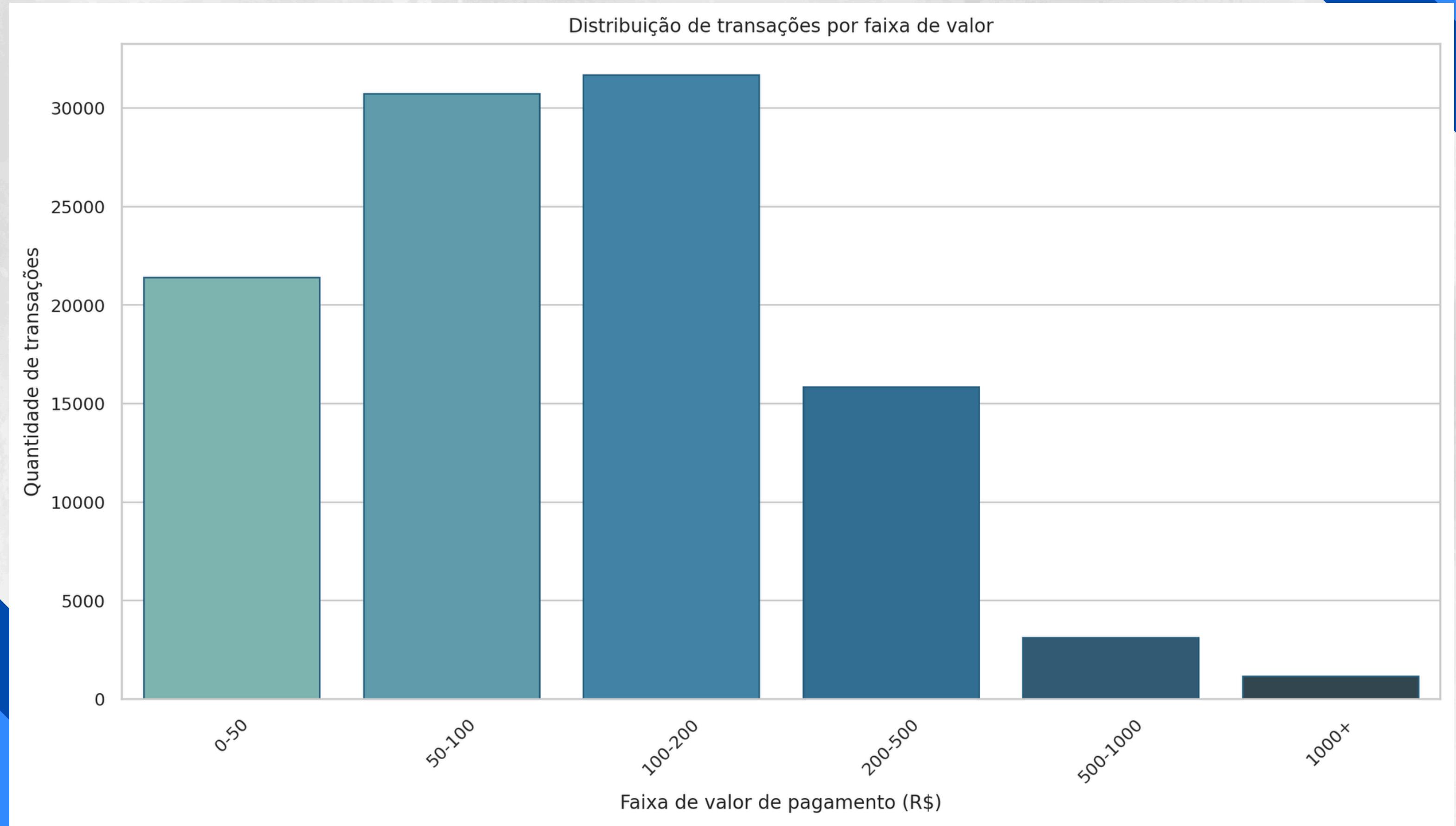
Análise sobre os meios de pagamento



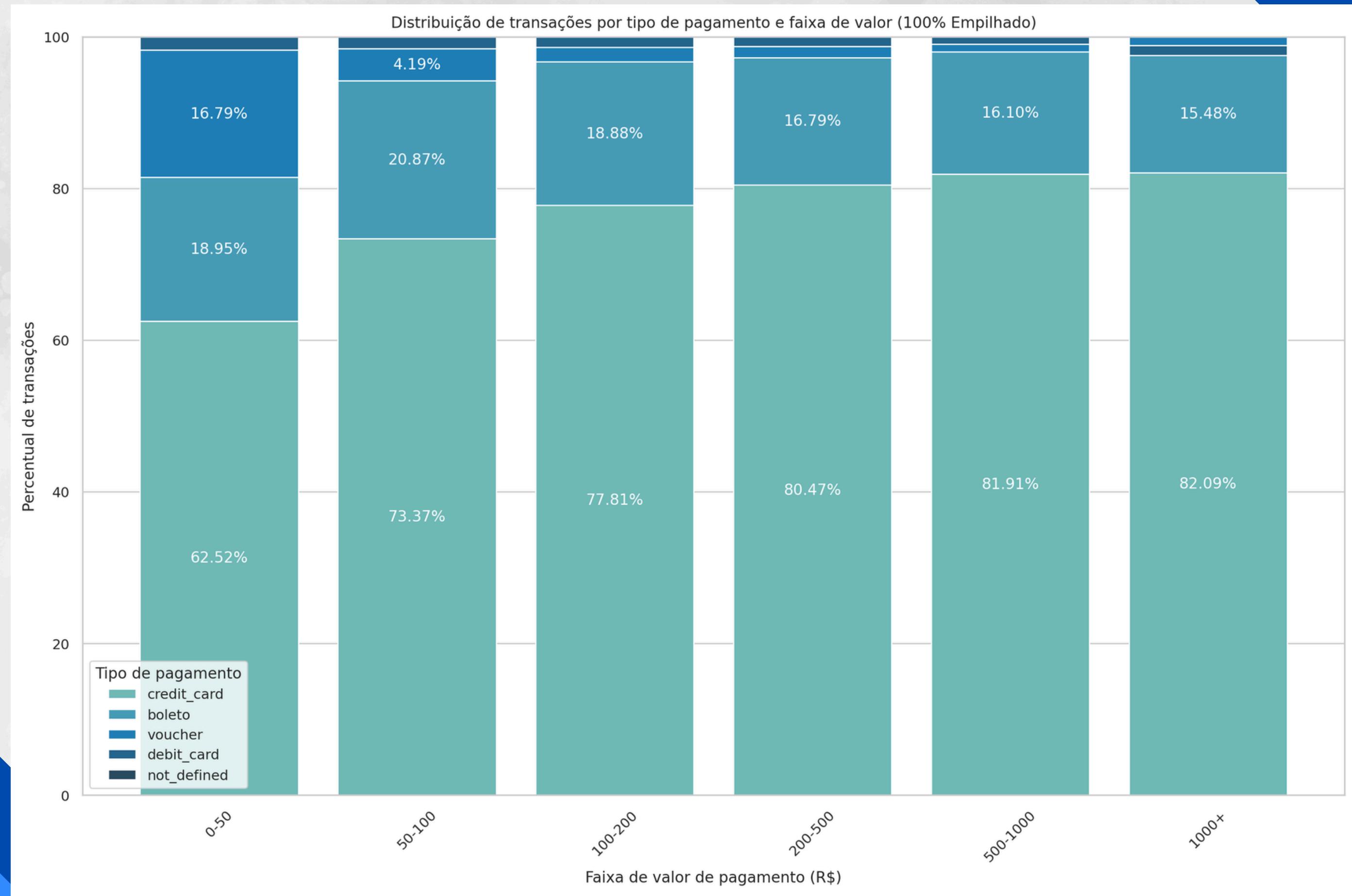
Análise sobre os meios de pagamento



Análise sobre os meios de pagamento



Análise sobre os meios de pagamento



CONCLUSÕES



CARTÃO DE CRÉDITO

- Presente na maioria das transações
- Aparece menos em transações de valores baixos



BOLETO

- Boa proporção em todas as faixas de valor
- Aparece menos em valores mais altos, talvez indicando dificuldade em pagar sem parcelas



VOUCHER

- Aparece bem nas faixas mais baixas, mas diminui bastante a partir de R\$ 100,00. Isso pode indicar falta de cupons que cubram transações altas



CARTÃO DE DÉBITO

- Nunca representa mais do que 2% de uma faixa, indicando pouco uso no e-commerce
- É similar ao boleto, aparecendo menos em faixas altas
- Nas transações acima de R\$ 1000,00, o cartão de débito aparece pela primeira vez maior do que o voucher

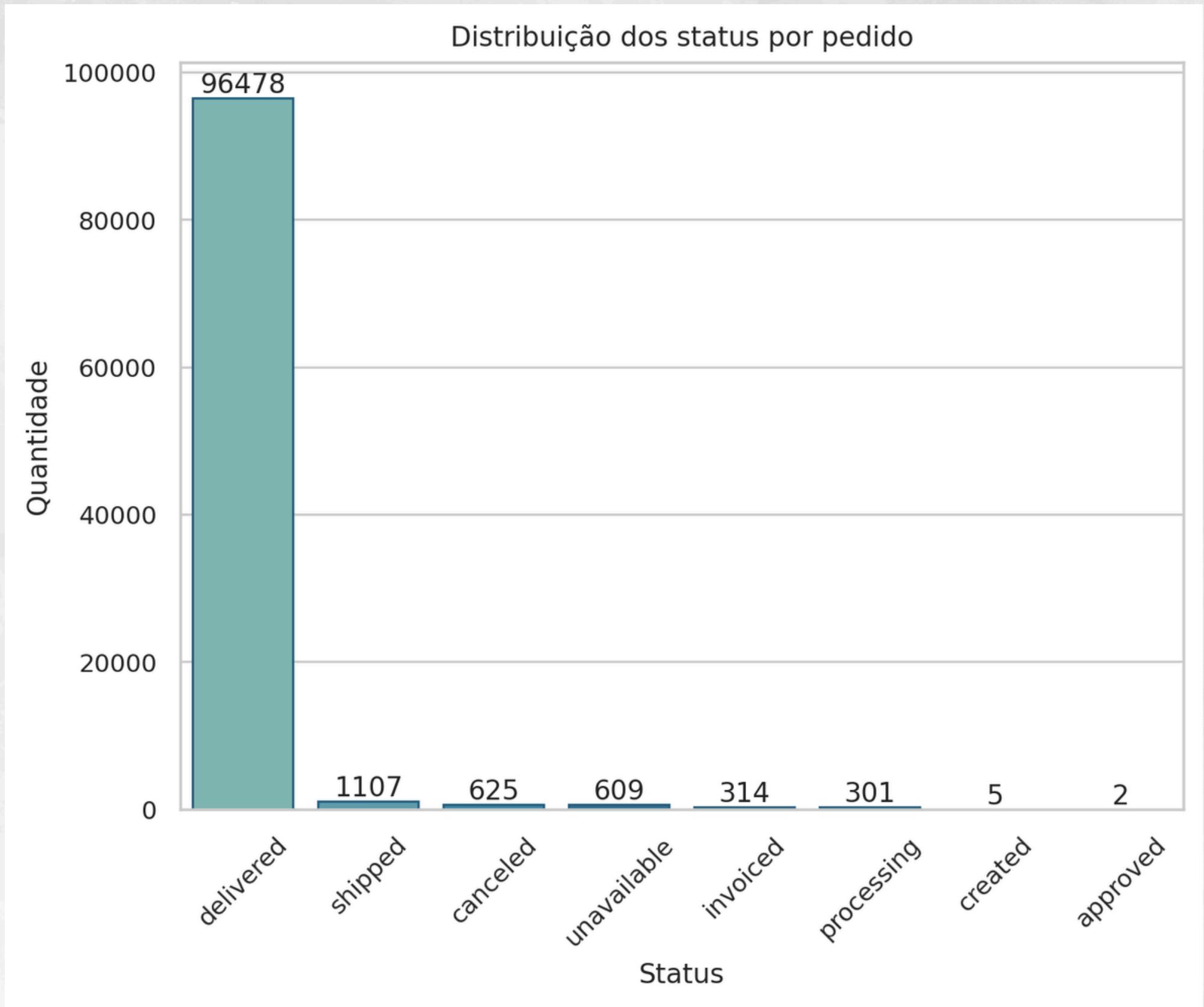


MEIO DESCONHECIDO

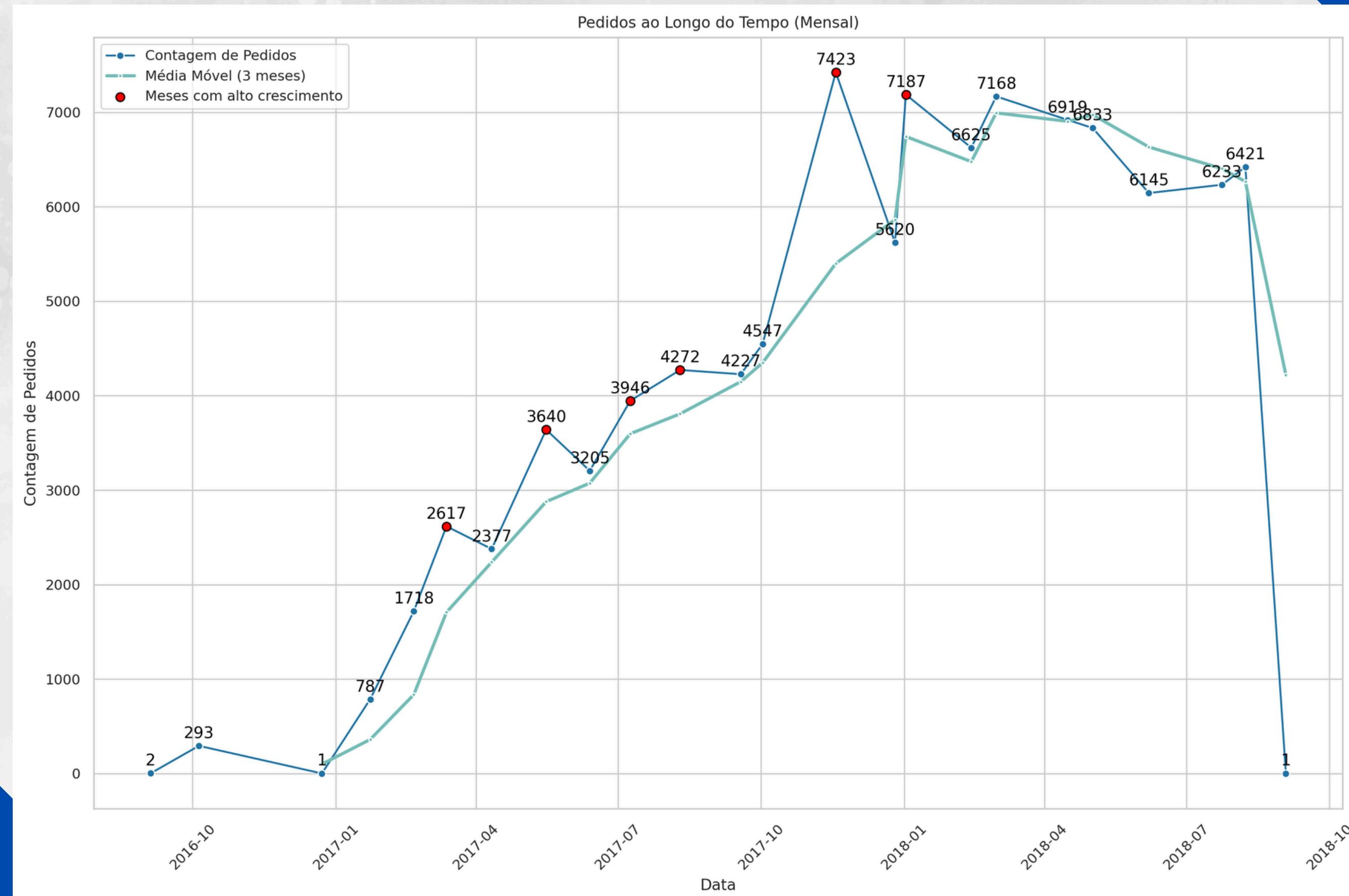
- Aparece em transações sem custo
- Pode indicar uma doação, amostra grátis, etc.



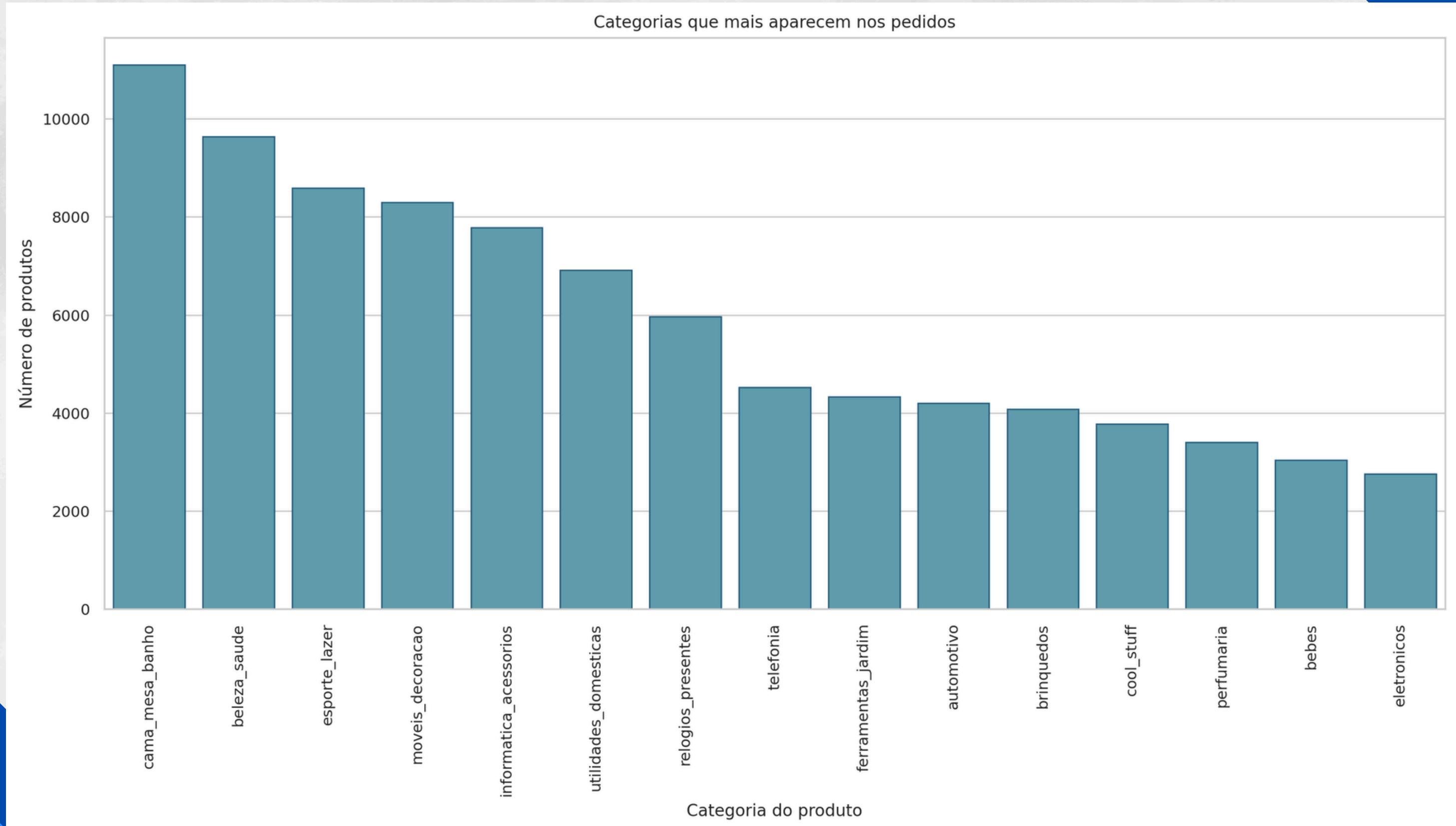
Análise sobre os pedidos - status



Análise sobre os pedidos - quantidade

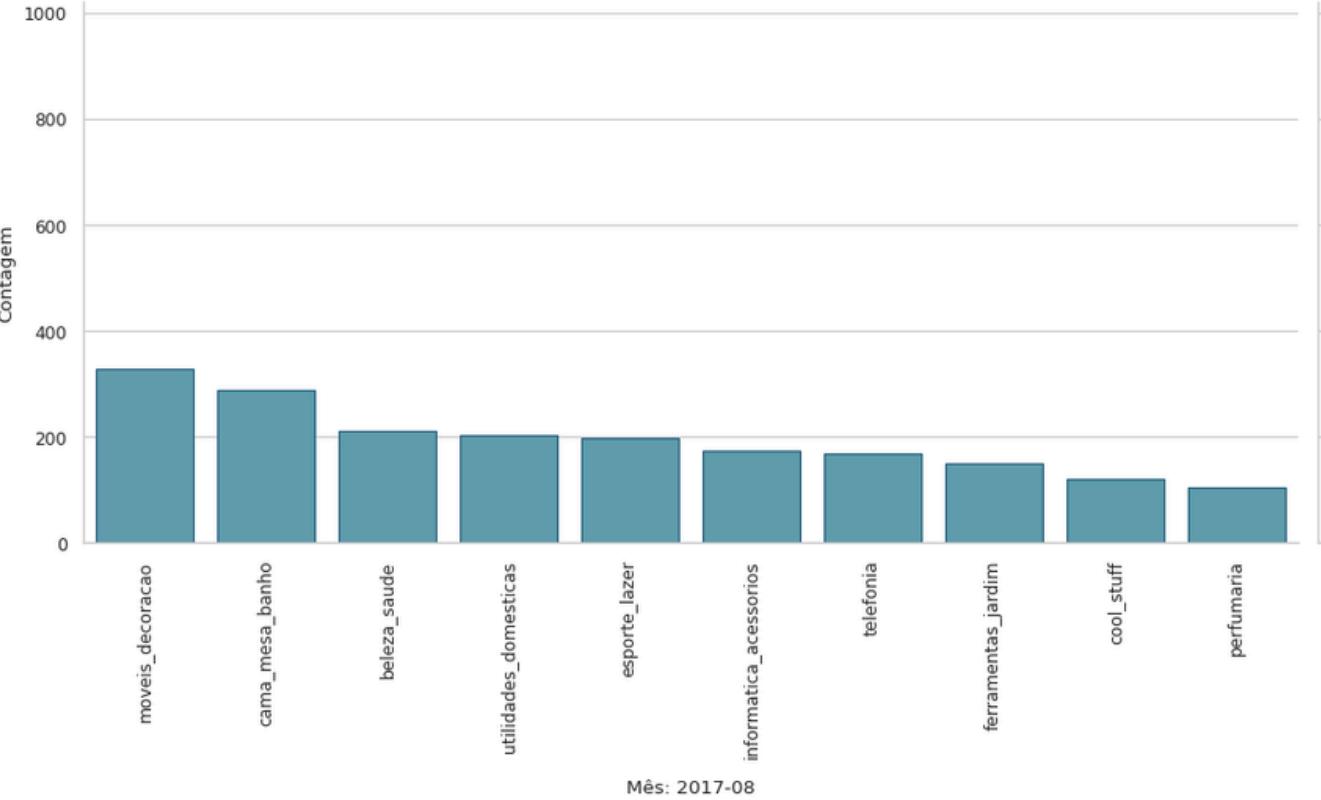


Análise sobre os pedidos - consumo geral

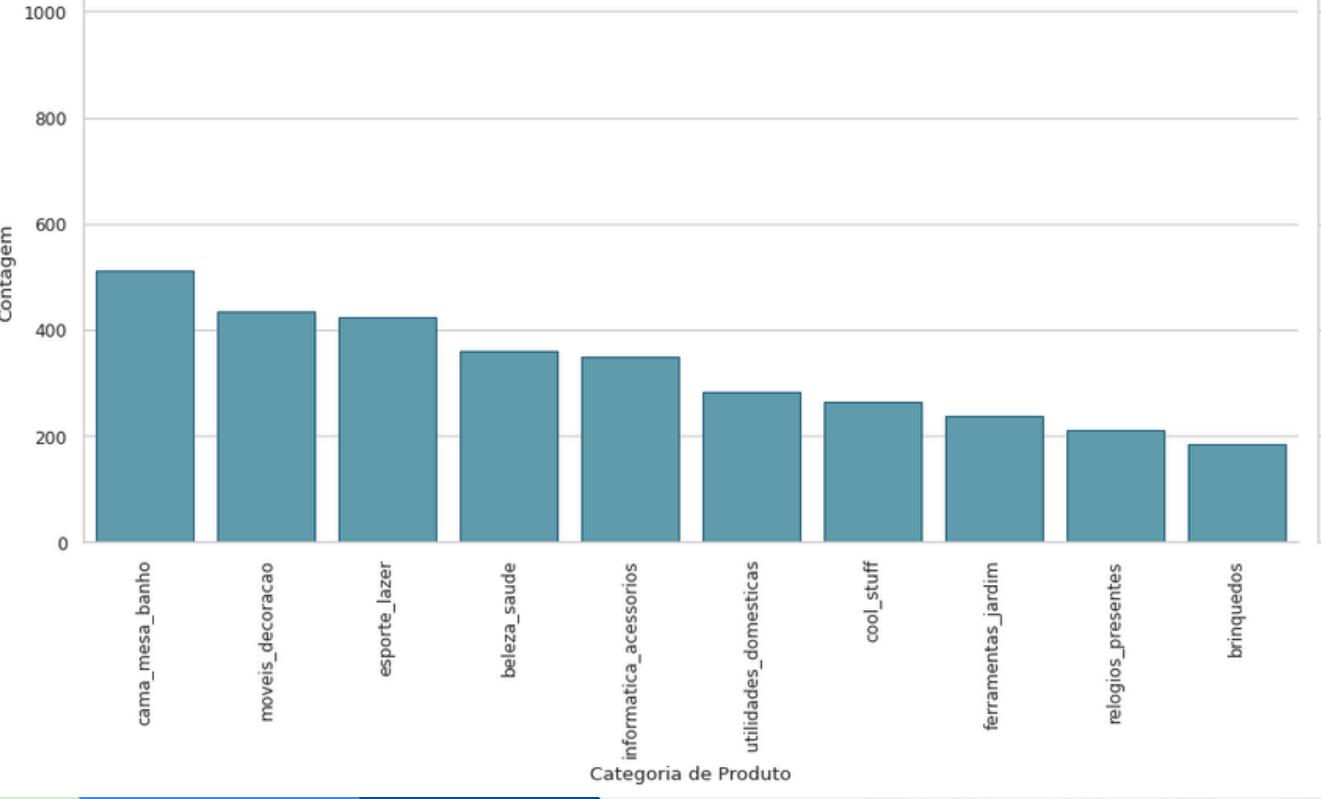


Análise sobre os pedidos - meses de maior variação

Mês: 2017-03

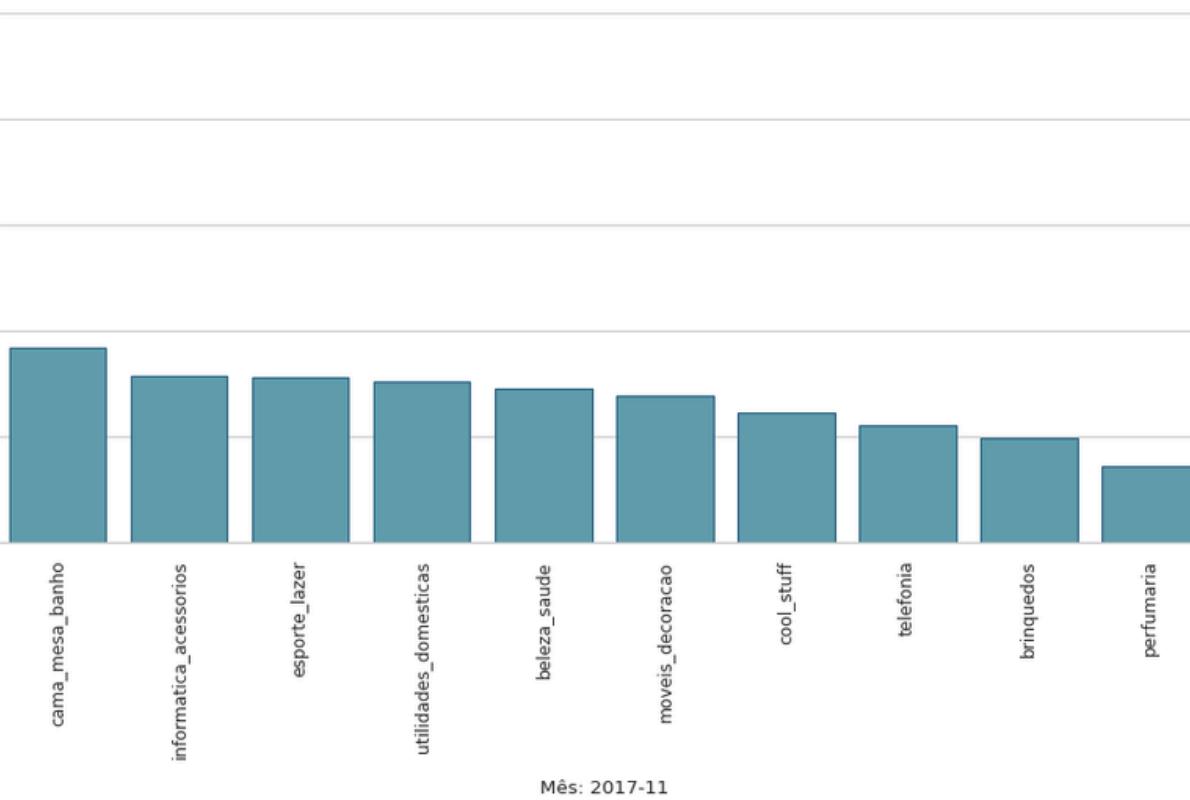


Mês: 2017-08

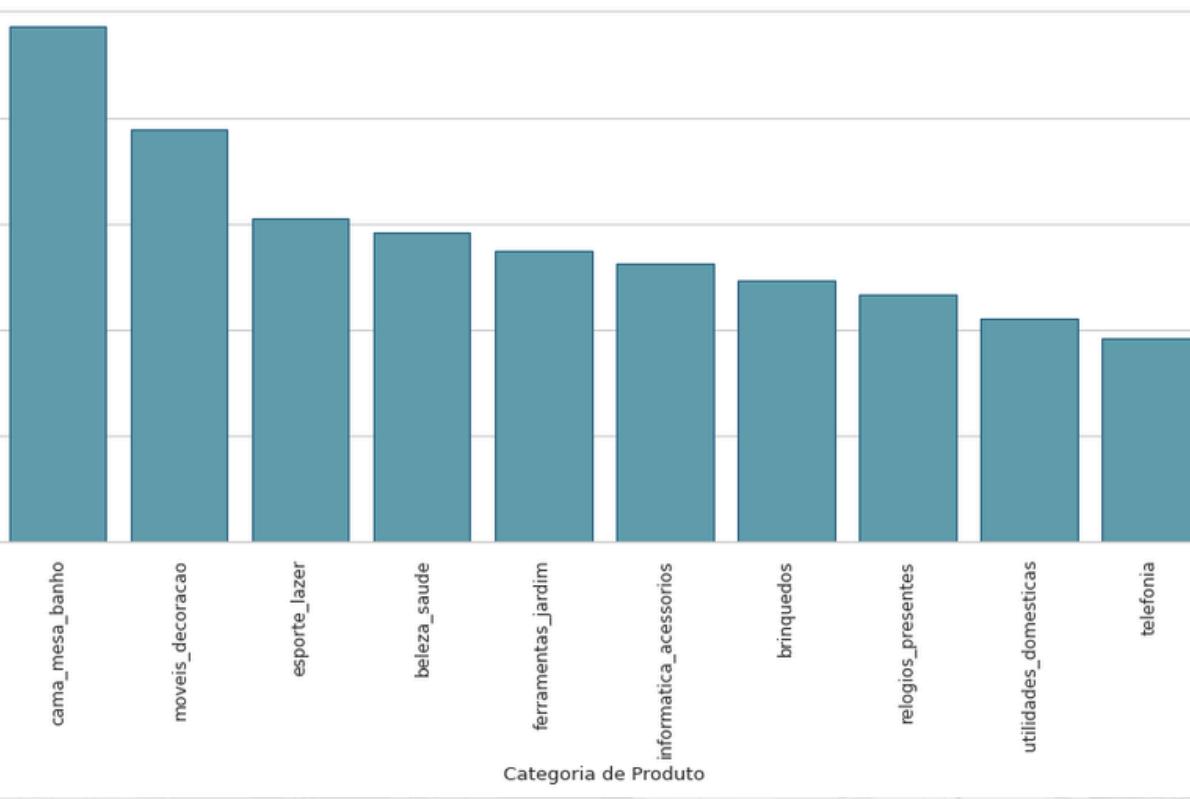


Produtos mais comprados nos meses de maior variação

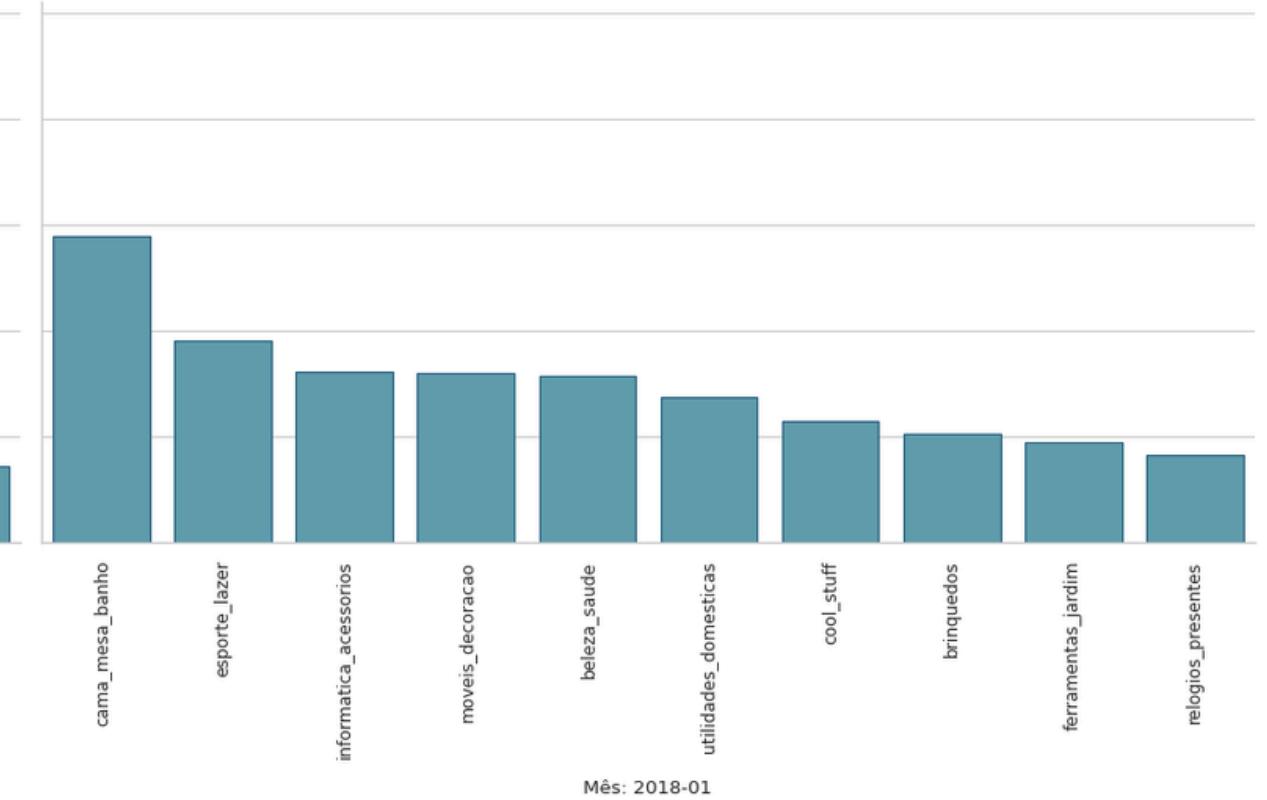
Mês: 2017-05



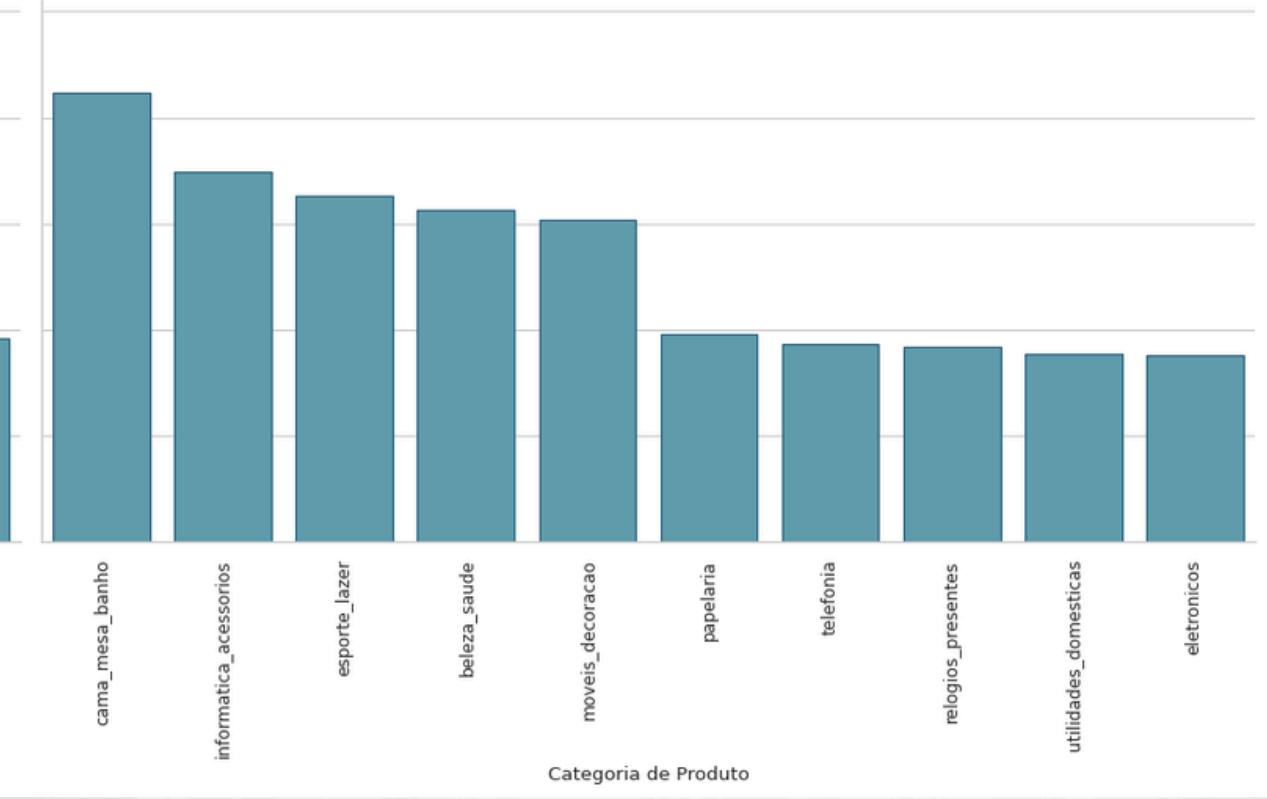
Mês: 2017-11



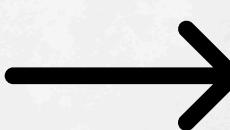
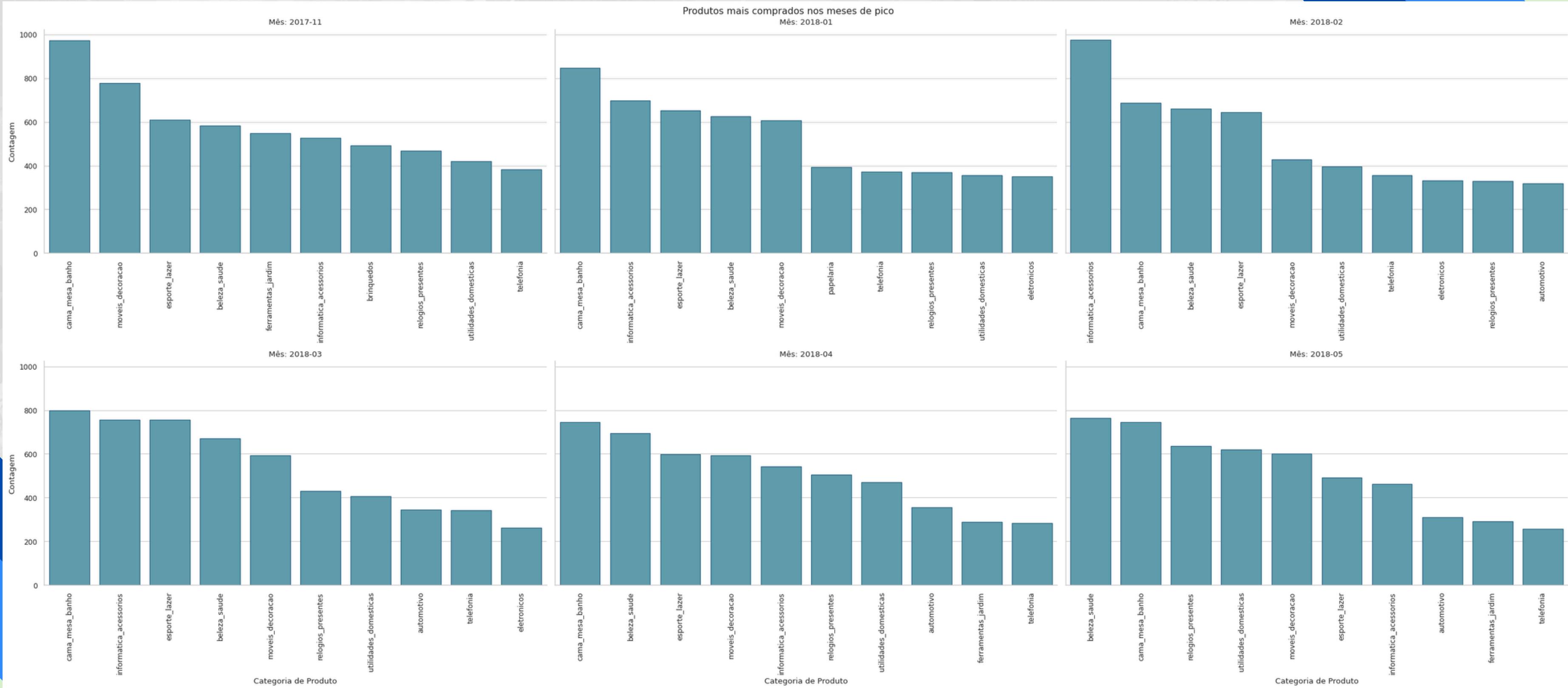
Mês: 2017-07



Mês: 2018-01



Análise sobre os pedidos - meses de maior variação



CONCLUSÕES



MESES COM MAIOR VARIAÇÃO

Geralmente são meses com algum evento sazonal:

- **Novembro, mês de Black Friday e Ciber Monday:**
 - Mostra ser um mês com promoções em diversas categorias
 - Categorias de produtos mais caros aparecem aqui, indicando uma certa tendência a comprar produtos mais caros neste mês - possivelmente por conta dos descontos
- **Para janeiro, mês de volta às aulas e liquidações de ano novo:**
 - Há um aumento na categoria 'papelaria', devido a volta às aulas e inicio do ano letivo
 - Categorias similares as de novembro, provavelmente por conta das liquidações de ano novo
- **Em março ocorre o Dia do Consumidor:**
 - Além dos itens de sempre, algumas outras categorias aparecem aqui, como 'cool_stuff', 'ferramentas_jardim' e 'perfumaria'
- **Em maio, mês de Dia das Mães:**
 - Vemos categorias com uma distribuição bem equilibrada
 - Temos 'beleza e saúde', 'acessorios de informática', 'utilidades domésticas' como algumas dessas categorias, provável tendência nos presentes para as mães
- **Agosto, mês do Dia dos Pais:**
 - Além das categorias mais frequentes, vemos outras categorias subindo, como 'ferramentas_jardim', 'perfumaria' e 'relogios_presentes', indicando, assim como em maio, uma provável tendência de presentes para os pais
- **Julho não possui um evento de grande destaque além de liquidações de inverno**



VISÃO GERAL DOS PEDIDOS

Produtos que aparecem ao longo dos meses

- No geral, o brasileiro possui uma alta em produtos das categorias de 'cama, mesa e banho', 'beleza e saúde', 'esporte e lazer' e 'decoração'
- Em 2018, seguindo de janeiro até maio, temos cinco dos seis meses com maiores números de pedidos
 - Como não temos datas especiais para cada um dos meses, não há tanto sentido associar isso diretamente a eventos sazonais
 - Talvez isso mostre um aumento na popularidade do e-commerce brasileiro em 2018, ou um mercado aquecido neste ano



Análise de Vendas

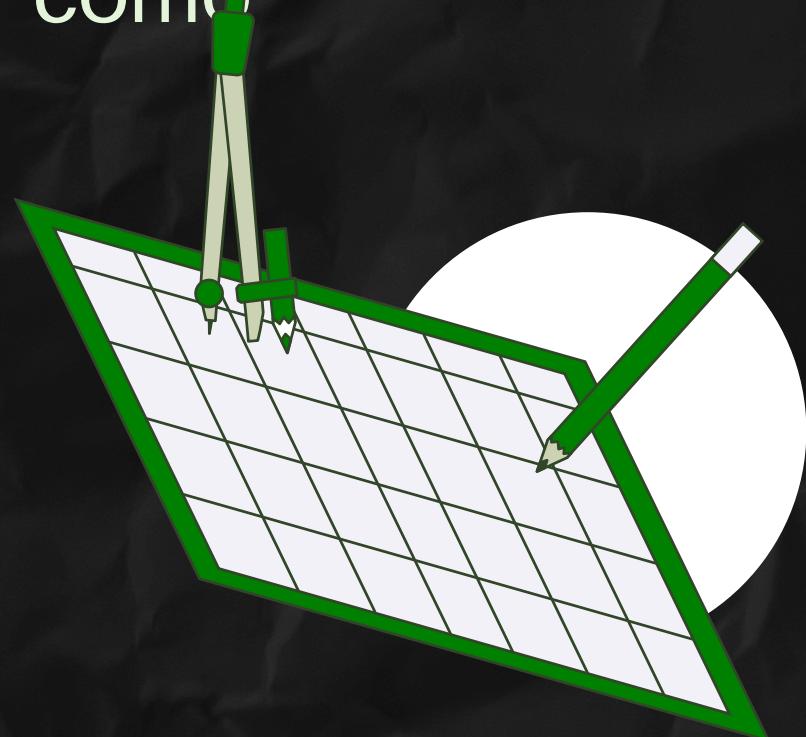
- Vendas por Produto e Categoria:
 - No geral, o brasileiro possui uma alta nas categorias de 'cama, mesa e banho', 'beleza e saúde', 'esporte e lazer' e 'decoração'.
 - Padrão que se repete ao longo do ano.
- Análise de Descontos e Ofertas:
 - Black Friday, Ciber Monday: Padrão se repete, mas com categorias de produtos geralmente mais caros ganhando destaque.
 - volta às aulas e liquidações de ano novo: aumento de vendas na categoria papelaria
 - Dia do Consumidor em março: 'cool_stuff', 'ferramentas_jardim' e 'perfumaria' aparecem
 - Dia das Mães: 'beleza e saúde', 'acessorios de informática', 'utilidades domésticas', mas com distribuição equilibrada
 - Dia dos Pais: 'ferramentas_jardim', 'perfumaria' e 'relogios_presentes'

Análise de Vendas

- Nos primeiros meses de 2018, é notável um crescimento de 'informatica_acessorios', conseguindo superar a categoria 'cama, mesa e banho' no mês de fevereiro
- Vemos uma sequência em 2018, seguindo de janeiro até maio.
 - Como não temos datas especiais para cada um dos meses, não há tanto sentido associar isso diretamente a eventos sazonais.
 - Popularização do E-commerce no período pré-pandemia.

Análise de Vendedores

- Performance de Vendedores:
 - Os vendedores que estão localizados nos estados de SP, PR e MG são os que mais realizam vendas na plataforma.
 - A variação desses vendedores oscila a medida que a quantidade de vendas também sofre de determinadas sazonalidades, como apresentado.



Análise de Usuários e Dados Geográficos

- Distribuição Geográfica dos Clientes
 - SP, RJ e MG dominam na distribuição das UF's, e os 3 que menos consomem são AP, AC e AM.
 - Região sudeste é líder em compras, que pode ser associado a maior quantidade de residentes e promoções (ex: frete grátis), gerando maior faturamento.
 - Região Norte gerando um menor faturamento e percentual de participação.

Análise de Pedidos e Avaliações

- **Análise de Pagamentos:**

- O método de cartão de crédito é o mais comum, de maneira geral, seguido por boleto, perdendo destaque a medida que o valor da compra vai aumentando.
- Débito e voucher são pouco representativos em todas as faixas de preço analisados.
- Voucher ganha destaque significativo em pedidos menores de 100 reais.
- Débito ganha vantagem em cima do voucher na primeira vez somente em compras acima de 100 reais.

OBRIGADO!
PERGUNTAS?

THANK
YOU
FOR
YOUR
PATIENCE
AND
THANK