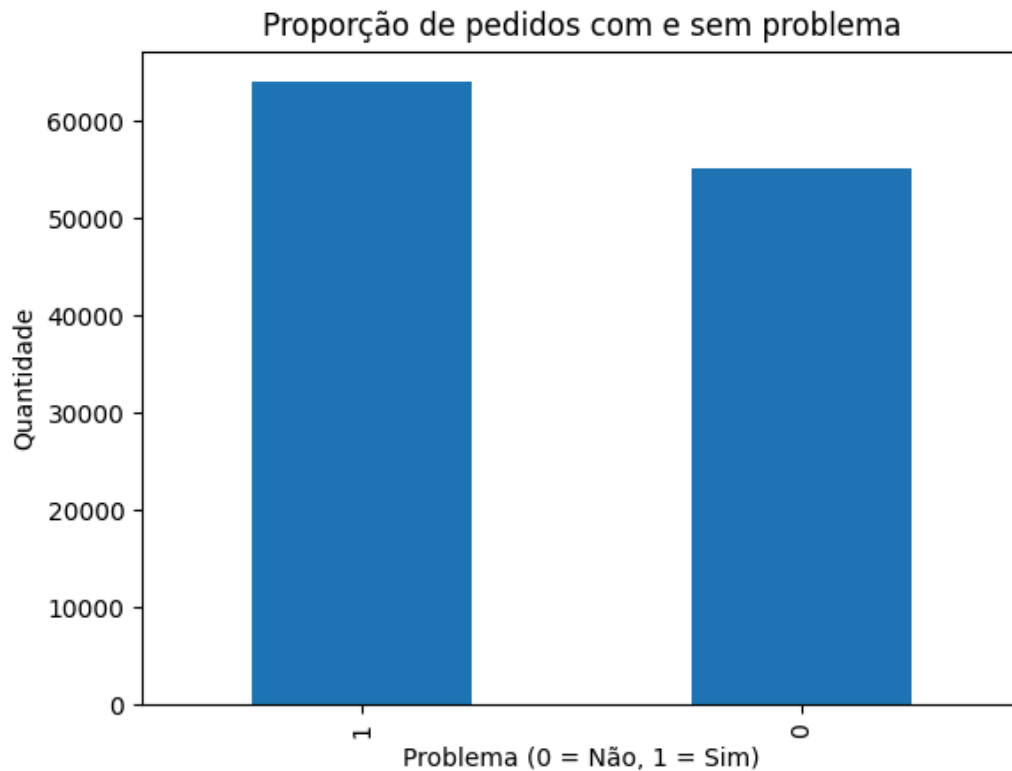


Análise Exploratória de Dados (EDA)

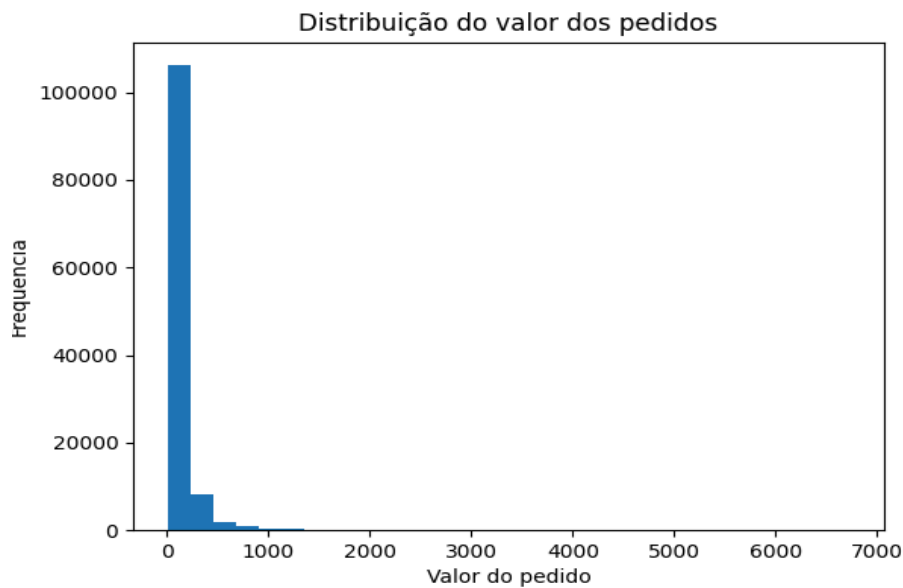
- ◆ Parte 1 – Distribuições Básicas

1. Qual a proporção de pedidos com problema vs. pedidos normais?



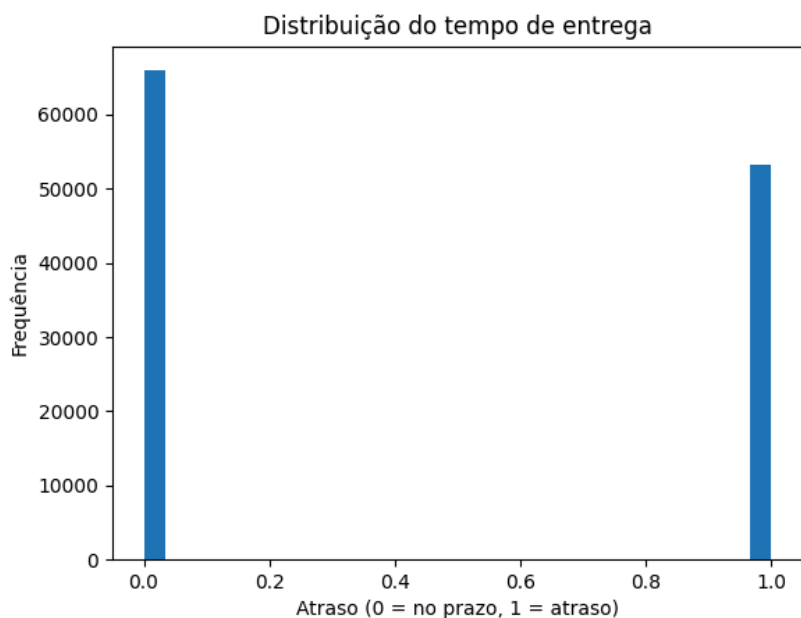
O gráfico de barras mostra que a maioria dos pedidos não apresenta problema, embora exista uma parcela relevante com ocorrência de problemas. Isso indica que, apesar do processo funcionar bem na maior parte do tempo, ainda há espaço para melhorias operacionais.

2. Como se distribui o valor dos pedidos?



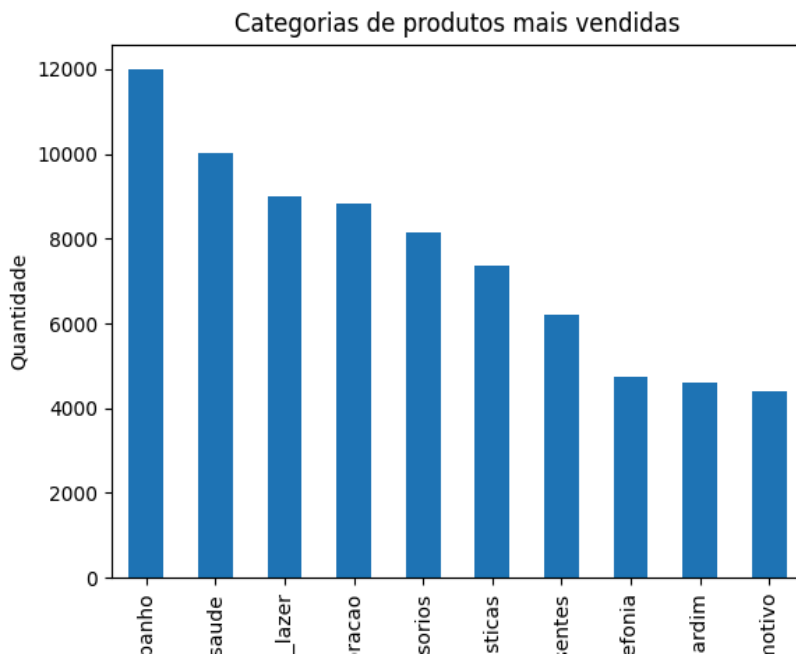
O histograma indica uma distribuição assimétrica à direita, com grande concentração de pedidos de baixo valor. Pedidos muito altos existem, mas são raros e caracterizam outliers.

3. Como se distribui o tempo de entrega?



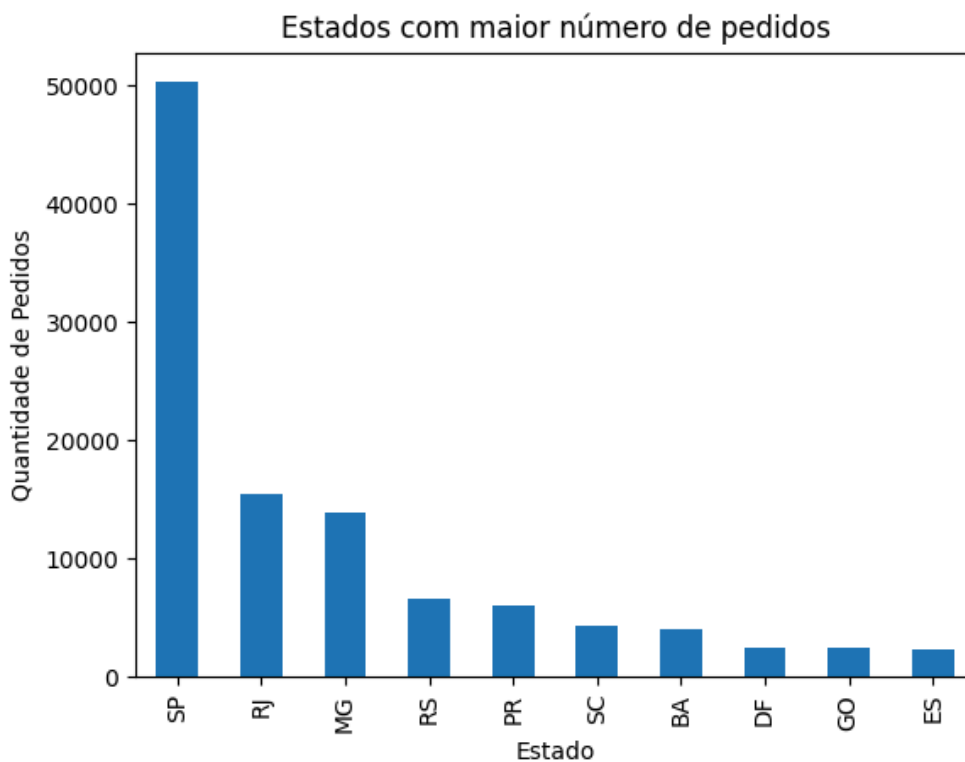
A distribuição é majoritariamente concentrada em entregas sem atraso, enquanto os atrasos aparecem como uma parcela menor. Isso sugere que atrasos não são a regra, mas impactam uma quantidade significativa de pedidos.

4. Quais são as categorias de produtos mais vendidas?



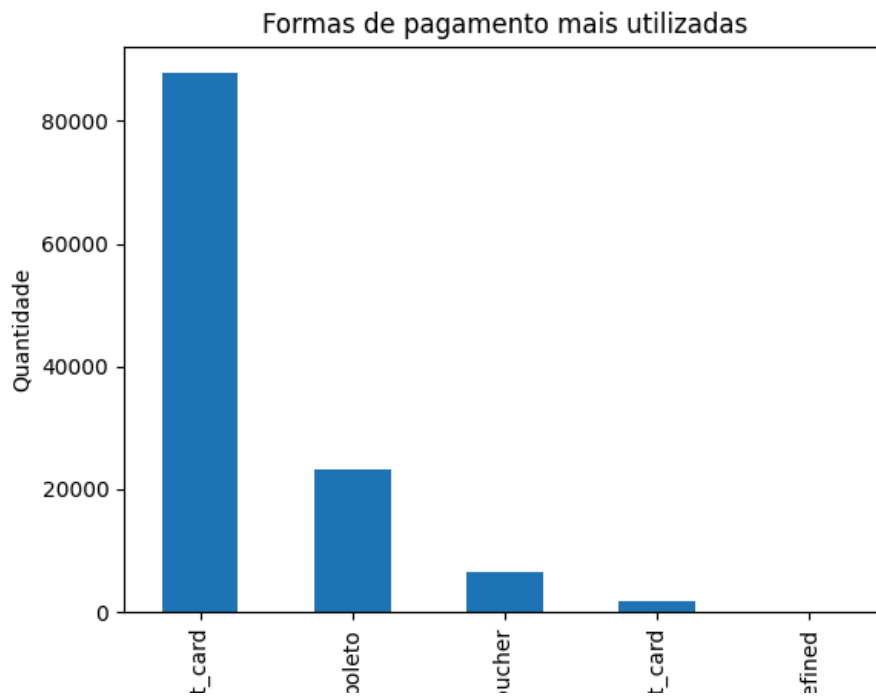
O gráfico de barras mostra que cama mesa, banho, beleza saúde e esporte lazer lideram em volume de vendas. Essas categorias representam a maior parte da demanda da plataforma.

5. Quais são os estados com maior número de pedidos?



São Paulo (SP) lidera com folga o número de pedidos, seguido por RJ e MG. Isso reflete a concentração populacional e econômica nesses estados.

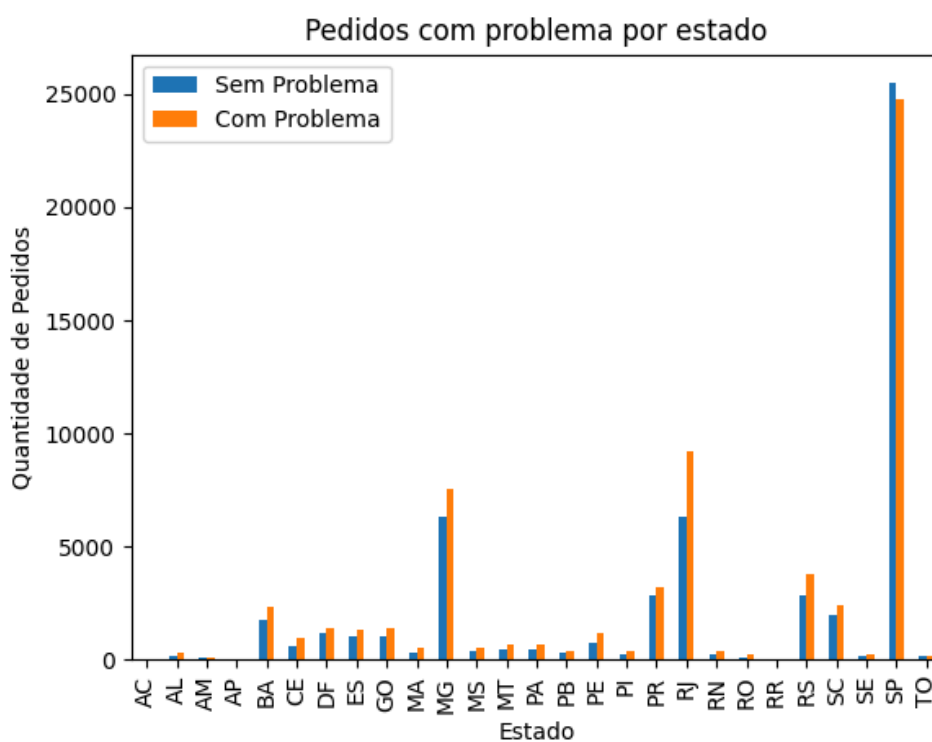
6. Quais são as formas de pagamento mais utilizadas?



O cartão de crédito é, disparadamente, a forma de pagamento mais utilizada. Outras formas como boleto e voucher aparecem com menor frequência, enquanto débito é pouco usado.

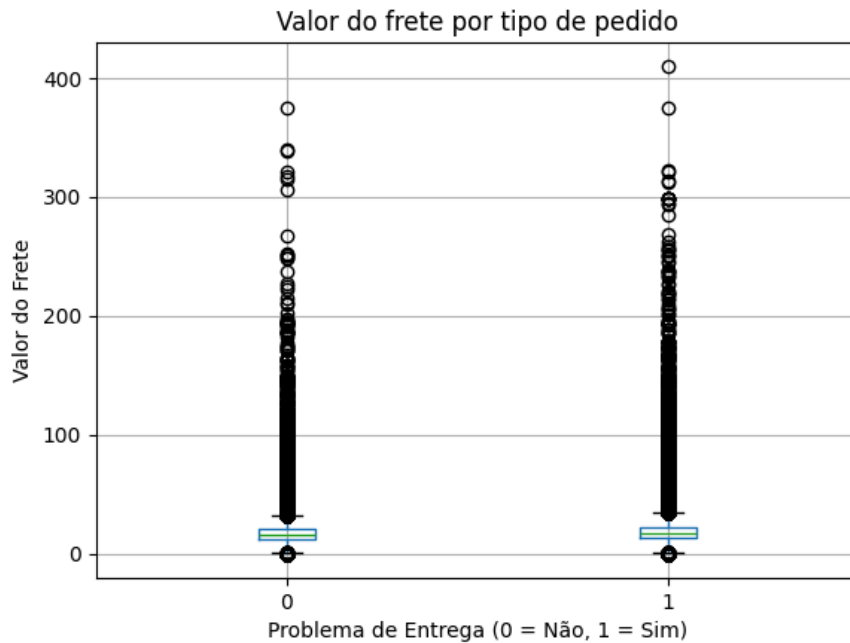
♦ Parte 2 – Análises Avançadas

1. Pedidos com problema variam por estado?



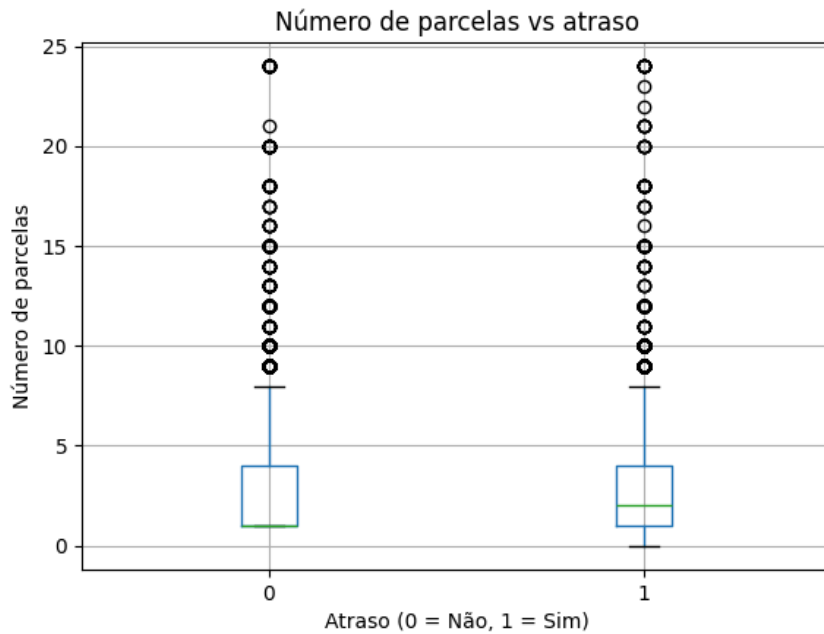
O gráfico indica que estados com maior volume de pedidos também concentram mais problemas. No entanto, proporcionalmente, alguns estados apresentam uma taxa de problemas relativamente maior.

2. O valor do frete influencia a chance de problema?



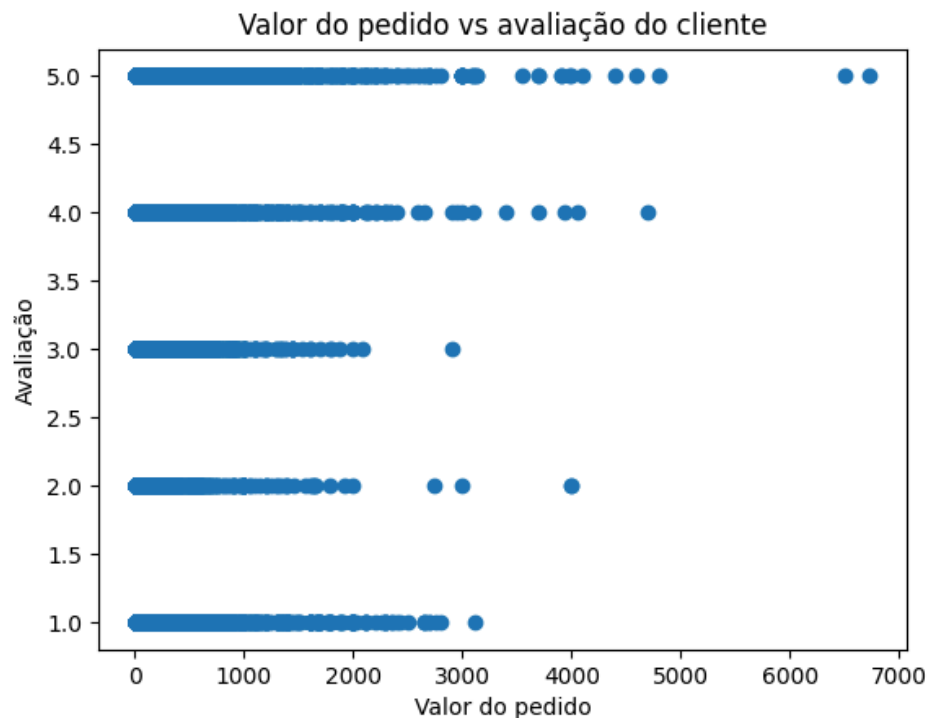
O boxplot mostra que pedidos com problema tendem a apresentar fretes ligeiramente mais altos. Isso sugere que distâncias maiores ou logística mais complexa aumentam o risco de falhas.

3. O número de parcelas influencia a chance de atraso?



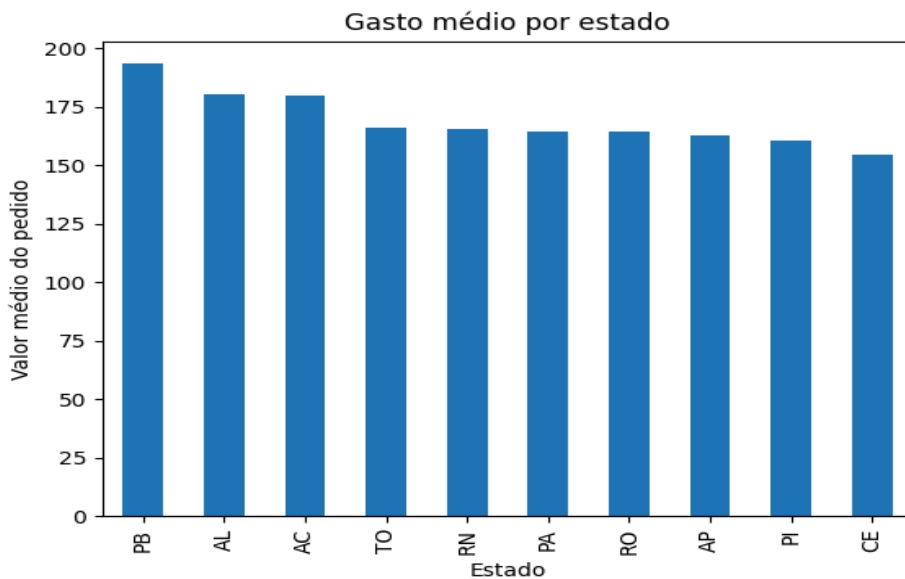
Pedidos atrasados apresentam uma mediana ligeiramente maior de parcelas. Isso pode indicar que compras mais caras ou planejadas tendem a ter maior risco de atraso.

4. Existe relação entre valor do pedido e avaliação do cliente?



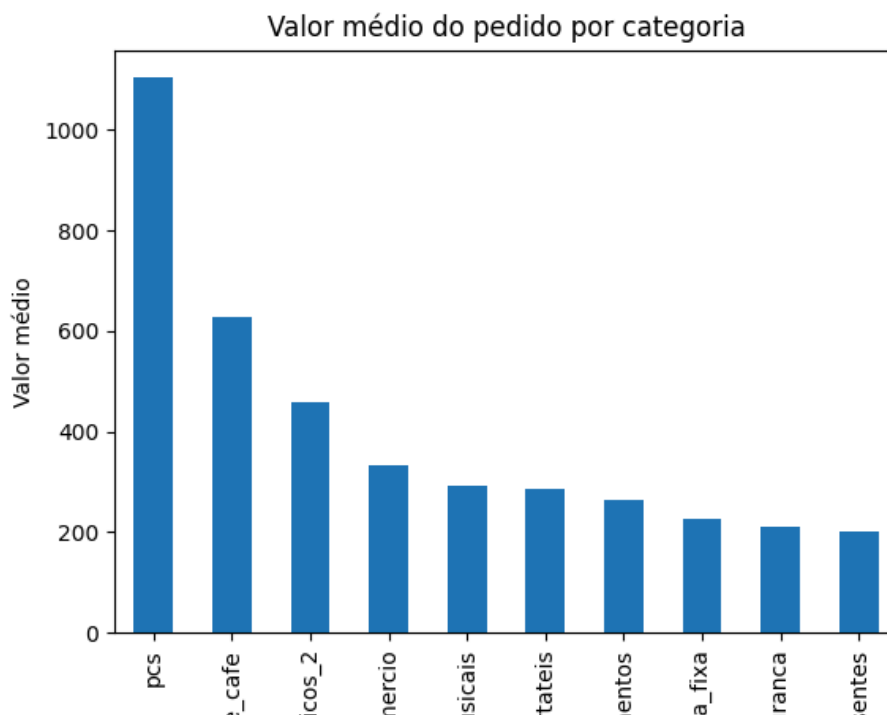
O gráfico de dispersão não mostra uma correlação forte. Avaliações altas e baixas aparecem em praticamente todas as faixas de preço, indicando que o valor não é o principal fator de satisfação.

5. Clientes de determinados estados gastam mais em média?



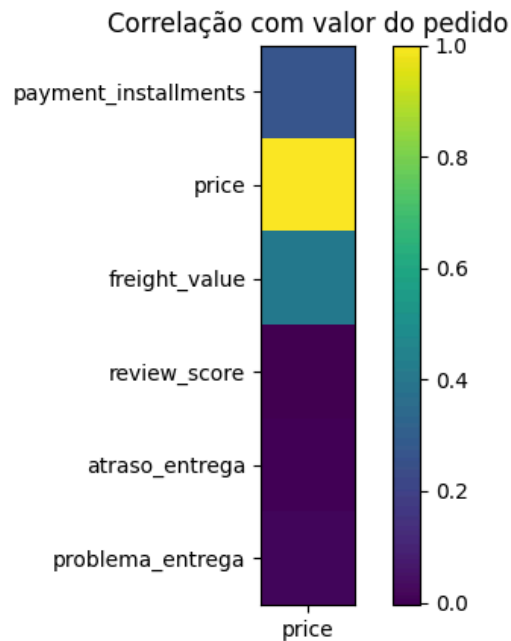
Estados como PB, AL e AC apresentam maior valor médio por pedido. Isso indica que, embora tenham menor volume de compras, os pedidos tendem a ser mais caros.

6. Há diferença no valor médio por categoria de produto?



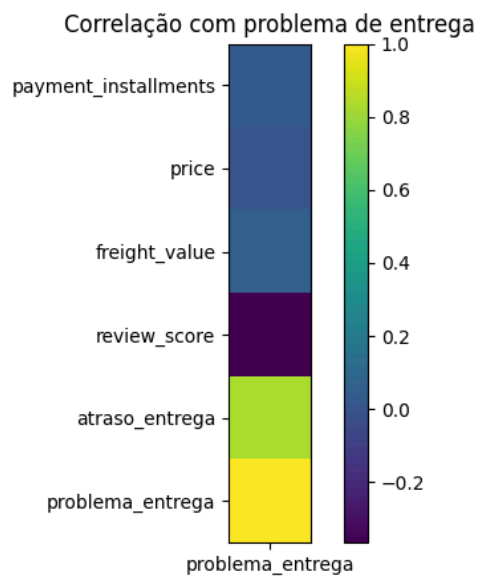
Categorias como pcs, portáteis casa_forno_e_cafe e eletrodomésticos possuem valores médios mais elevados. Já as categorias populares têm maior volume, mas menor ticket médio.

7. Quais variáveis têm maior correlação com o valor do pedido?



O heatmap mostra forte correlação entre valor do pedido e valor do frete, além de uma correlação moderada com o número de parcelas. Avaliação e problemas têm pouca relação com o valor.

8. Quais variáveis têm maior correlação com PROBLEMA?



A variável mais correlacionada com problema é o atraso na entrega, indicando forte relação entre atraso e ocorrência de problemas. Outras variáveis apresentam correlação fraca ou irrelevante.