

Relatório de Análise de Regressão – PicMoney

A presente análise tem como objetivo aplicar técnicas de regressão linear simples para explorar relações entre variáveis de interesse da empresa PicMoney, a partir de três bases distintas: **Transações**, **Massa de Teste** e **Pedestres**. O intuito é compreender em que medida determinadas variáveis explicativas podem prever os valores de saída observados nas bases, avaliando tanto a força quanto a direção das associações, a significância estatística dos coeficientes estimados e a qualidade geral do ajuste do modelo.

O processamento dos dados foi realizado em linguagem **R**, utilizando pacotes como *dplyr*, *ggplot2*, *broom* e *tidyr*. Inicialmente, os valores numéricos passaram por uma etapa de limpeza por meio da função `limpar_numerico()`, que padronizou formatos e eliminou ruídos textuais, garantindo consistência nos cálculos. Em seguida, foram aplicados testes de correlação de Pearson para avaliar a intensidade e a direção da associação entre variáveis, além da construção de modelos de regressão linear simples. Para melhor visualização, foram extraídas amostras aleatórias de 200 observações em cada base, permitindo a construção de gráficos de dispersão com a reta de regressão ajustada, o que facilita a interpretação visual dos resultados obtidos.

Base de Transações

Na base de transações, investigou-se a relação entre o **valor do cupom** e o **repassé à PicMoney**. O teste de correlação de Pearson indicou um coeficiente de 0,369, sinalizando uma associação positiva de intensidade moderada. O valor-p associado ao teste foi menor que $2,2e-16$, o que permite rejeitar a hipótese nula de ausência de correlação, confirmando que existe uma relação estatisticamente significativa entre as variáveis.

O modelo de regressão linear simples ajustado apresentou a seguinte equação:

$$\text{repassé_picmoney} = -0,64 + 0,129 * \text{valor_cupom}$$

O coeficiente angular de 0,129 indica que, a cada aumento unitário no valor do cupom, espera-se, em média, um acréscimo de 0,129 no valor do repasse à PicMoney. O intercepto, embora não significativo ($p = 0,304$), sugere que para valores próximos de zero no cupom, o repasse tende a ser nulo ou próximo disso.

A qualidade do ajuste, avaliada pelo R^2 , foi de 0,136, ou seja, aproximadamente 13,6% da variabilidade do repasse é explicada pelo valor do cupom. Apesar de ser um resultado significativo do ponto de vista estatístico, o poder explicativo do

modelo é limitado, o que indica que existem outros fatores além do valor do cupom que influenciam o repasse.

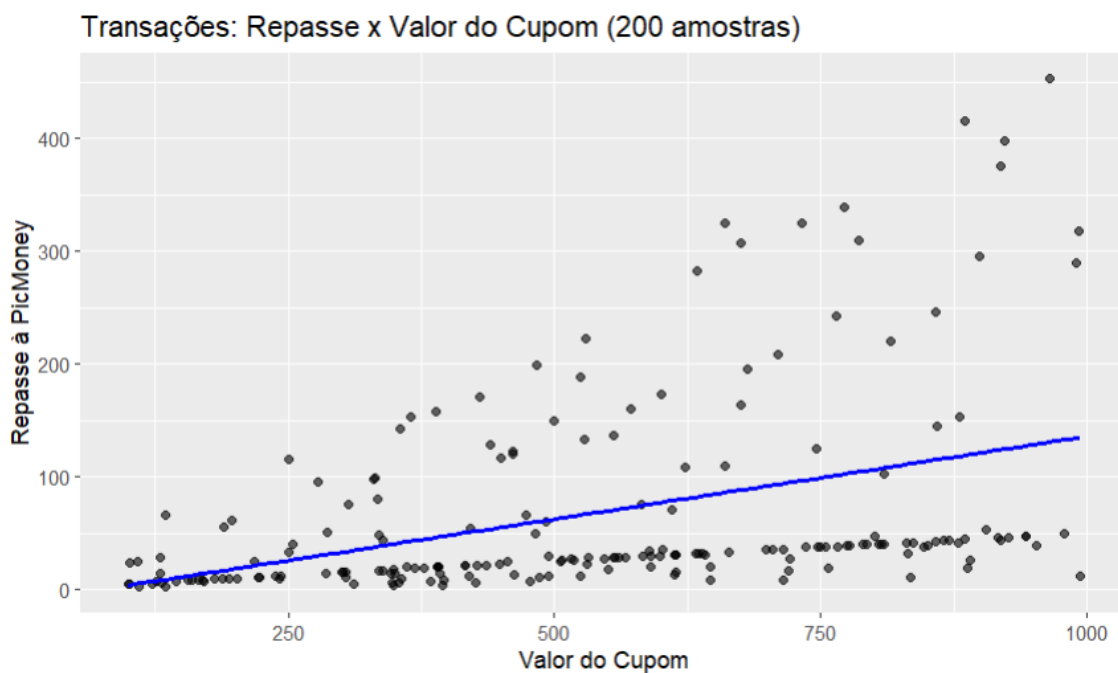
Quando a análise é segmentada por **tipo de cupom**, o cenário se altera substancialmente. Para cupons do tipo **Cashback**, a correlação foi praticamente perfeita ($r \approx 1,0$), com slope igual a 0,05. Isso significa que o repasse é diretamente proporcional ao valor do cupom, em uma relação quase determinística. Para os cupons de **Desconto**, a correlação foi de 0,743, e a equação estimada foi:

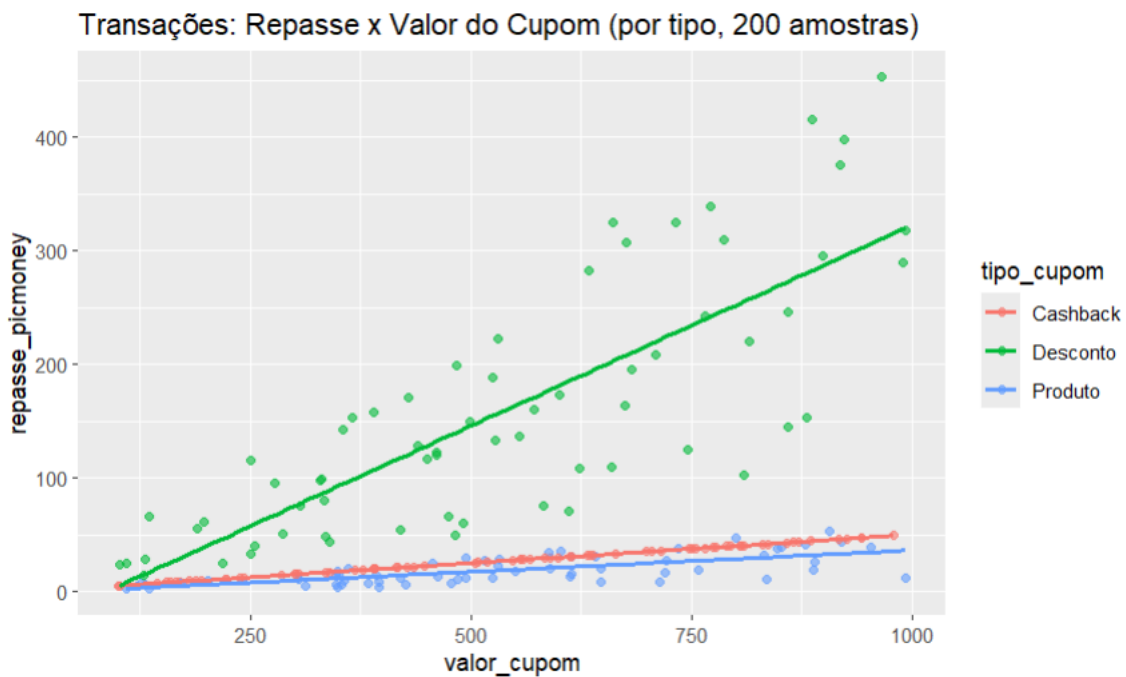
$$\text{repasse_picmoney} = -0,211 + 0,300 * \text{valor_cupom}$$

Já para os cupons do tipo **Produto**, a correlação foi de 0,723, e a equação foi:

$$\text{repasse_picmoney} = -0,0877 + 0,0353 * \text{valor_cupom}$$

Esses dois últimos casos apresentam associações fortes, com inclinações positivas, o que reforça que, independentemente do tipo, valores maiores de cupom levam a maiores repasses.





Base de Massa de Teste

Na base de massa de teste, o interesse foi investigar a relação entre o **valor da compra** e o **valor do cupom**. A correlação geral entre as variáveis foi de 0,375, semelhante à observada na base de transações: positiva, de intensidade moderada e altamente significativa ($p < 2,2e-16$).

O modelo de regressão linear ajustado apresentou a seguinte equação:

$$\text{valor_cupom} = -1,58 + 0,131 * \text{valor_compra}$$

O coeficiente angular de 0,131 significa que, a cada unidade acrescida no valor da compra, o valor do cupom tende a aumentar, em média, em 0,131. O intercepto não foi significativo, o que indica que, para compras próximas de zero, o valor do cupom não segue um padrão confiável. O R^2 obtido foi de 0,140, o que novamente revela baixo poder explicativo do modelo geral: apenas 14% da variação nos valores de cupons pode ser atribuída ao valor da compra.

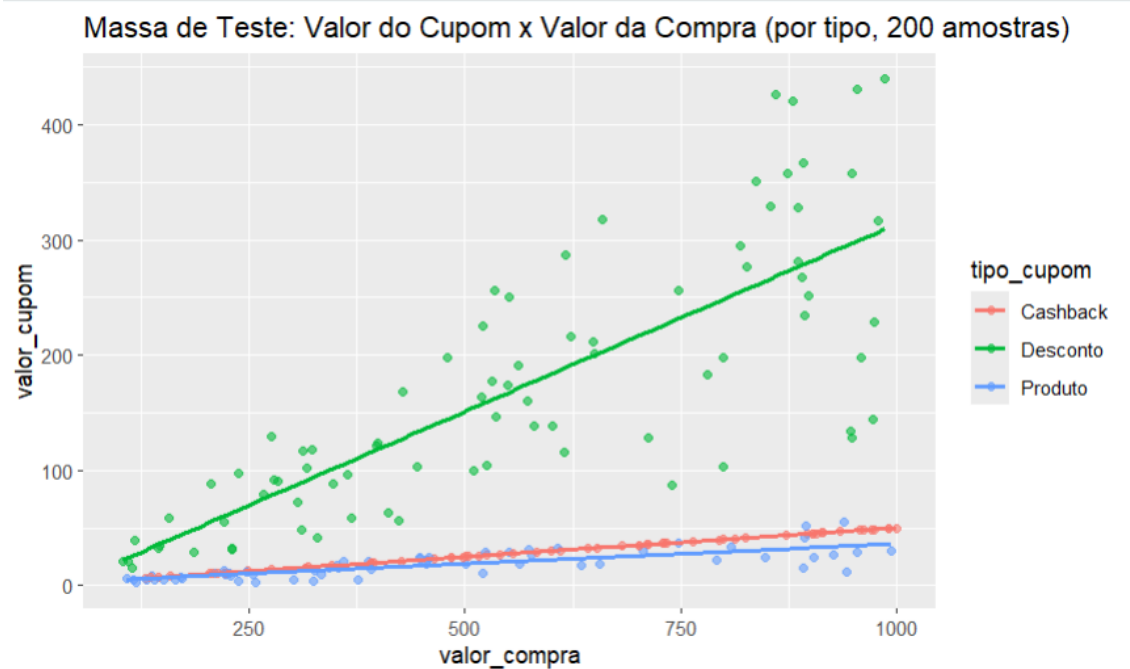
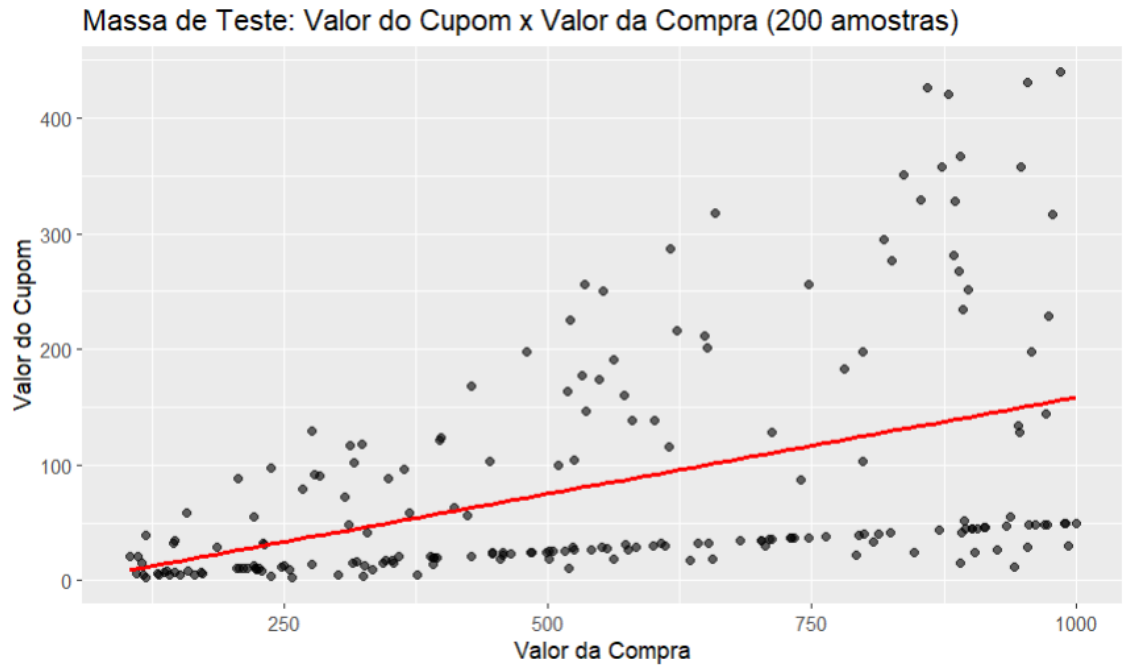
Assim como na base anterior, a análise segmentada por tipo de cupom forneceu modelos muito mais robustos. Para o tipo **Cashback**, observou-se uma correlação perfeita ($r = 1,0$), com slope igual a 0,05, caracterizando um modelo linear praticamente determinístico. Para o tipo **Desconto**, a correlação foi de 0,751, com equação:

$$\text{valor_cupom} = -2,72 + 0,304 * \text{valor_compra}$$

Já para o tipo **Produto**, a correlação foi de 0,732, com equação:

$$\text{valor_cupom} = -0,348 + 0,036 * \text{valor_compra}$$

Assim, a segmentação revelou que, embora a relação global seja fraca, dentro de cada categoria de cupom, especialmente Cashback, os modelos lineares são confiáveis e podem ser utilizados em previsões.



Base de Pedestres

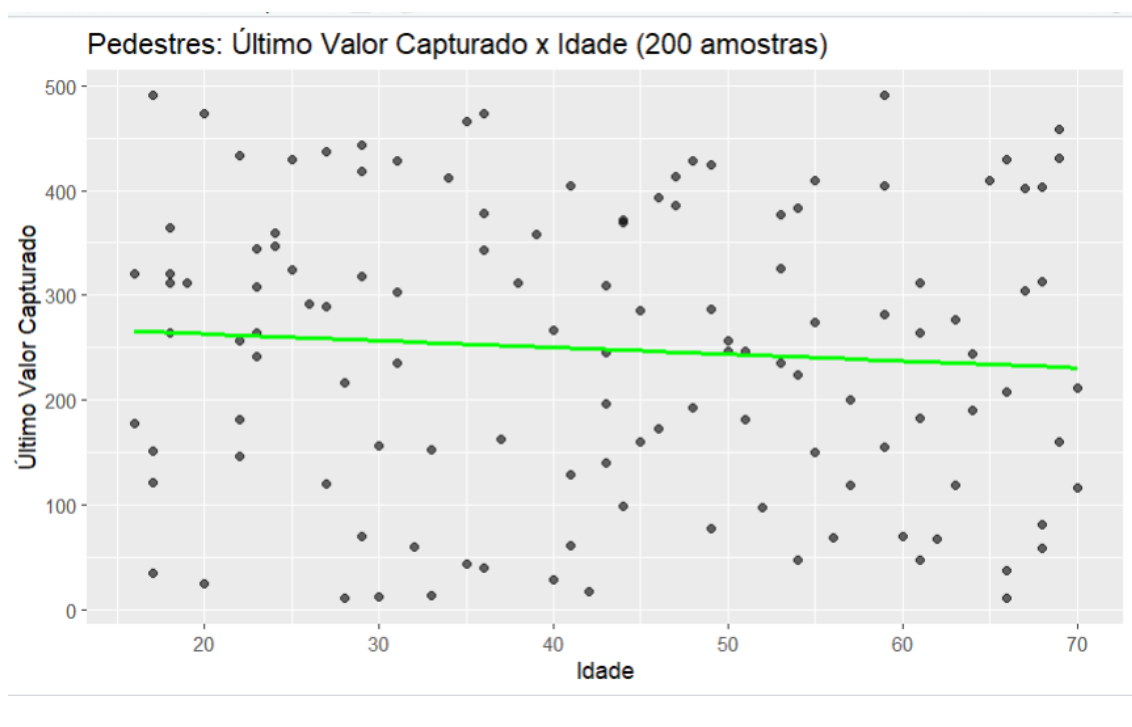
Na base de pedestres, a análise buscou identificar a relação entre a **idade** do indivíduo e o **último valor capturado**. A correlação encontrada foi de apenas

0,003, valor próximo de zero, com $p = 0,4561$. Isso significa que não há evidências de correlação entre as variáveis.

O modelo linear ajustado apresentou a equação:

$$\text{ultimo_valor_capturado} = 253,62 + 0,027 * \text{idade}$$

O coeficiente angular de 0,027 não foi estatisticamente significativo, indicando que variações na idade não têm impacto sobre o valor capturado. O R^2 foi próximo de zero ($9,26e-06$), o que confirma que o modelo não explica a variabilidade da variável dependente. Em outras palavras, a idade não é uma variável preditiva relevante neste contexto, e o modelo linear é incapaz de capturar qualquer relação significativa.



Conclusões

A análise de regressão realizada permitiu identificar diferentes padrões de associação entre variáveis em três bases distintas. Nas bases de **Transações** e **Massa de Teste**, as correlações gerais foram moderadas e os modelos apresentaram baixo poder explicativo (R^2 em torno de 0,14). Isso sugere que, quando consideradas de forma agregada, as variáveis **valor do cupom** e **valor da compra** não são bons preditores sozinhos.

No entanto, ao segmentar as análises por **tipo de cupom**, os resultados foram significativamente mais fortes. O tipo **Cashback** mostrou-se perfeitamente previsível, com correlação igual a 1 e modelos determinísticos. Já os tipos **Desconto** e **Produto** apresentaram correlações fortes (acima de 0,72), com

modelos que explicam de forma consistente a variação observada. Essa evidência indica que a segmentação é fundamental para construir modelos de regressão úteis para o negócio da PicMoney.

Por outro lado, a base de **Pedestres** demonstrou a irrelevância da idade como variável explicativa para o último valor capturado, uma vez que não houve correlação e o modelo linear ajustado não apresentou significância estatística nem poder explicativo.

Em síntese, as análises reforçam a importância de **considerar a natureza do cupom** para prever valores associados, destacando o tipo Cashback como determinístico e os tipos Desconto e Produto como bons candidatos a modelos preditivos. Já variáveis demográficas simples, como idade, não apresentam utilidade prática para modelagem nesse contexto.