

Nome/RA: Bernardo Seijas Cavalcante/24026290.

Nome/RA: Eduardo Chen Zou/24025817.

Nome/RA: Fabiano Henrique Chou/24025991.

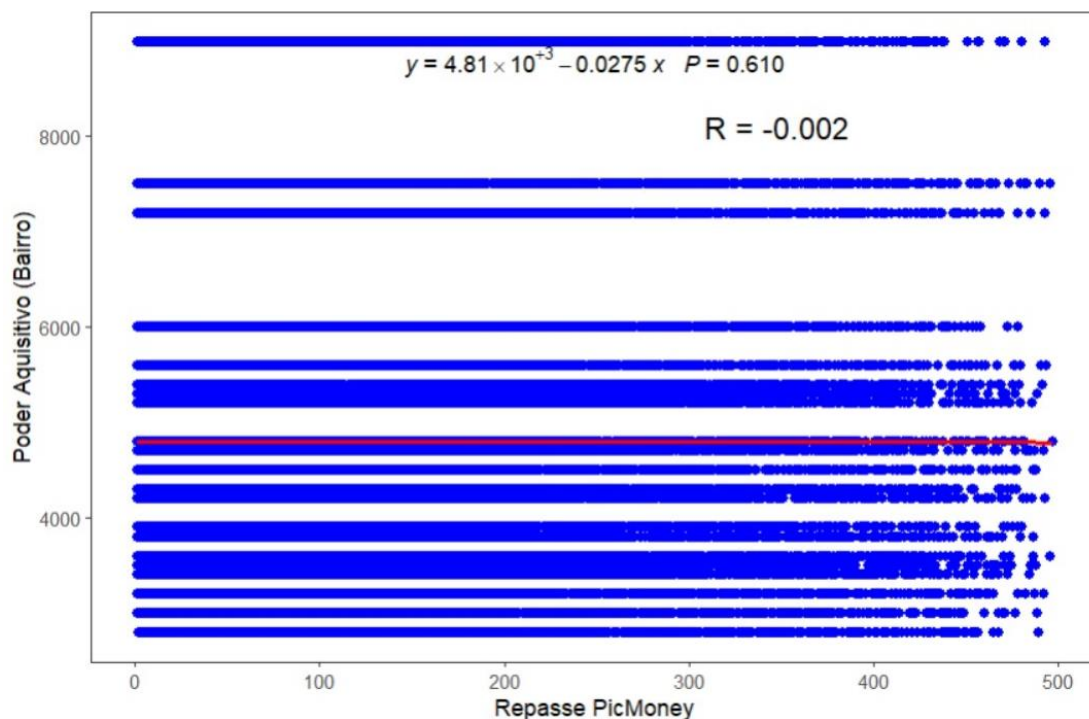
Nome/RA: Nicolas Morales/24025897.

22/09/2025

Ranking Bairros

O gráfico apresentado busca avaliar a relação entre o **poder aquisitivo dos bairros** (eixo Y) e o **repasse PicMoney** (eixo X). A análise foi realizada por meio de uma regressão linear, representada pela linha vermelha. O objetivo é identificar se existe dependência entre essas duas variáveis e se o poder aquisitivo de um bairro exerce influência nos repasses.

Os dados utilizados para a construção dos gráficos foram obtidos a partir de bases fornecidas pela PicMoney, IA e organizados em rankings. O ranking de **poder aquisitivo** corresponde a um índice que classifica os bairros segundo sua capacidade de consumo, em valores discretos previamente definidos. Já o ranking de **repasses** reflete os valores associados às transações realizadas pela plataforma. A partir dessas informações, foram gerados gráficos para verificar se existia algum tipo de relação entre as variáveis analisadas.



A reta ajustada apresentou a equação $y=4810-0,0275x$, a partir da qual conclui-se que os dados não possuem relação nenhuma entre si. Em cima

disso, depreende-se que mesmo que o alto poder aquisitivo de um bairro pudesse ter uma relação com alto valor de um cupom que, conseqüentemente, levaria a repasses com valores maiores, o gráfico demonstra que a IA que a PicMoney utilizou para gerar os dados definiu a mesma probabilidade de ter valores resultantes de repasse para todos os bairros independentemente de qualquer característica particular destes. Tal afirmação explica o porquê que cada faixa de poder aquisitivo é praticamente igual à outra. Nesse sentido, o Coeficiente de Correlação de Pearson é muito próximo de 0, implicando que a IA gerou uma cadeia de dados cujas relações a serem exploradas entre si não geram boas análises sobre as naturezas dos dados, como será reforçado nas posteriores análises que virão após esta.

1. Coeficiente de correlação (R):

O valor de **$R = -0,002$** indica **correlação nula**. Não existe tendência linear entre repasse e poder aquisitivo.

2. Coeficiente de determinação (R^2):

O valor de **R^2 é próximo de 0** (se $R = -0,002$, então $R^2 \approx 0,000004R$). Isso significa que o modelo linear explica praticamente **0% da variação** do poder aquisitivo com base nos repasses.

3. Significância estatística (P):

O valor de **$P = 0,610$** é muito maior que 0,05.

Isso indica que a inclinação da reta não é estatisticamente significativa — a associação encontrada é apenas fruto do acaso.

Conclusão

O gráfico demonstra que **não existe relação significativa entre o repasse PicMoney e o poder aquisitivo dos bairros**.

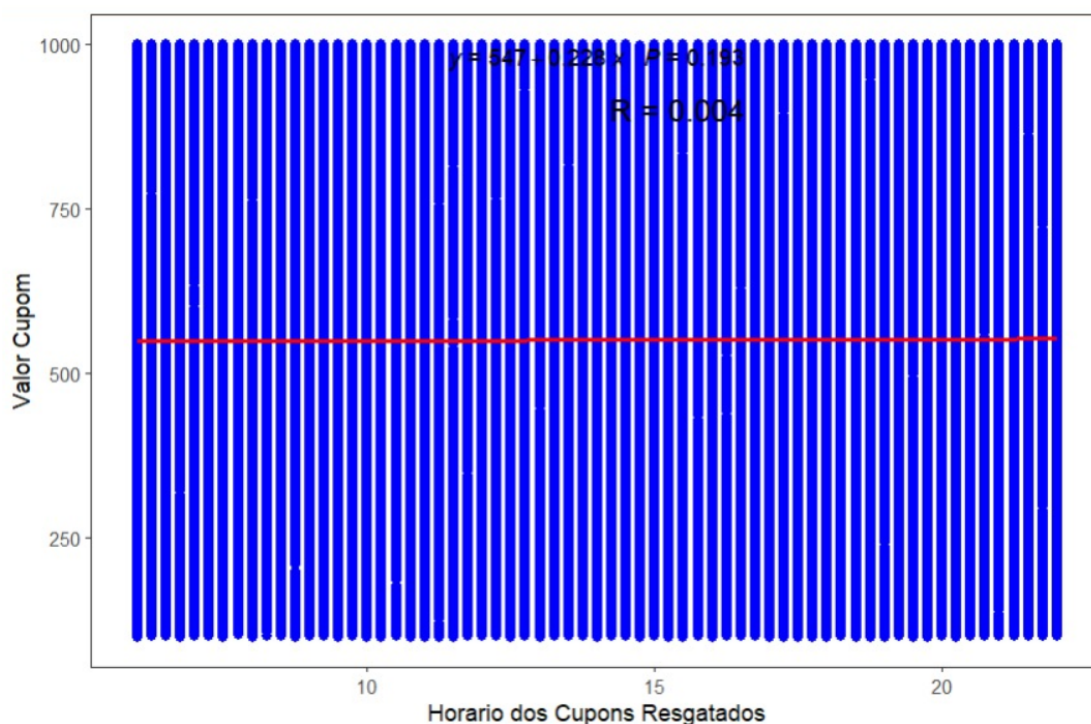
- O valor de **$R \approx 0$** mostra ausência de correlação.
- O **$R^2 \approx 0$** confirma que a regressão não explica a variabilidade dos dados.
- O **$P > 0,05$** comprova que o modelo não tem significância estatística.

Em resumo, **os repasses da PicMoney não dependem do poder aquisitivo**: bairros de maior ou menor renda apresentam volumes de repasse semelhantes, sem padrão claro ou tendência.

Horário dos Cupons que foram resgatados

A análise teve como objetivo verificar se existe alguma relação entre o horário em que os cupons foram resgatados e o valor de cada cupom. Para isso, foi elaborado um gráfico de dispersão contendo os pontos correspondentes aos valores dos cupons e o horário do dia convertido para um número racional de 0 a 24, acompanhados de uma reta de regressão linear ajustada aos dados. Essa abordagem permitiu avaliar visualmente se há tendência ou correlação entre as variáveis analisadas.

O ranking dos dados foi definido de forma automática pela IA, que distribuiu os valores de forma homogênea entre os horários, garantindo que todos os intervalos tivessem proporções semelhantes de cupons.



Além disso, o **Coefficiente de Correlação de Pearson**, que resultou, ao indicar ausência de correlação linear. Com base nisso, concluiu-se que, mesmo que o horário de resgate pudesse ter alguma influência no valor do cupom, a base de dados analisada mostrou que os cupons possuíam valores distribuídos de forma aleatória, sem qualquer padrão relacionado ao horário. Esse achado indicou que o modelo de geração dos dados da PicMoney atribuiu valores de cupons de forma uniforme ao longo do dia, independentemente de variações de horário.

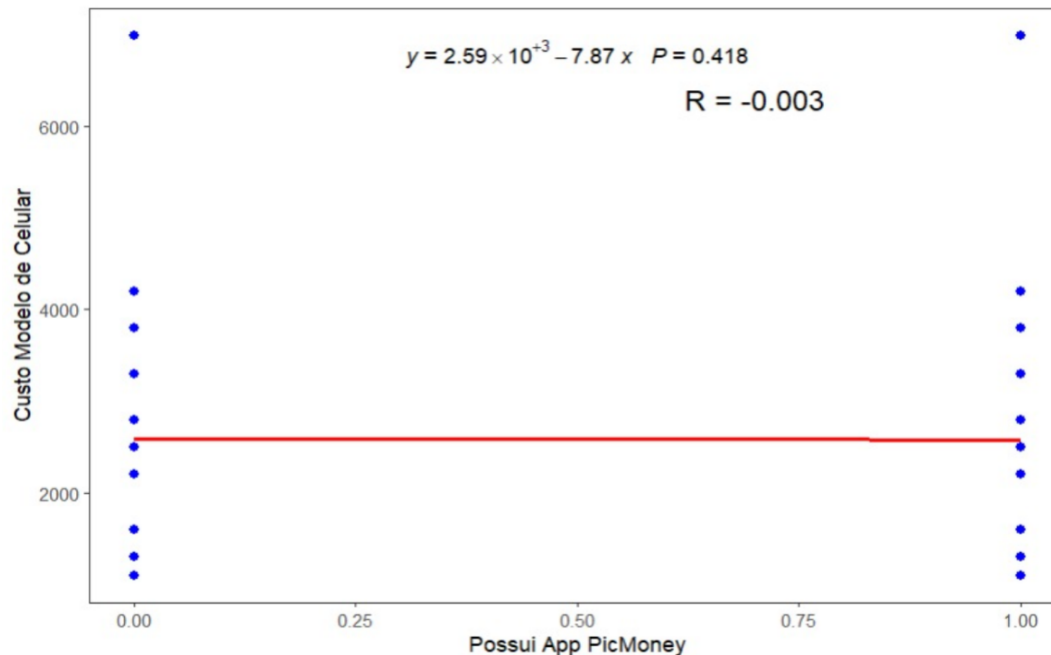
Conclusão

A reta ajustada apresentou a equação $y=541,06-0,004x$, o que indicou uma inclinação praticamente nula. O **Coefficiente de Correlação de Pearson** resultou em $R=0,004$, e o **coeficiente de determinação** (R^2) foi de aproximadamente 0,000016. Esse valor é extremamente próximo de zero, demonstrando que menos de 0,002% da variação no valor dos cupons pode ser explicada pelo horário em que foram resgatados — na prática, uma explicação nula.

Além disso, pela forma como os dados foram gerados, a distribuição se apresenta extremamente homogênea, fazendo com que o gráfico se assemelhe a uma série de linhas verticais. Isso reduz a utilidade da aplicação de regressão linear nesse caso, pois não há variabilidade suficiente para gerar uma análise interpretativa significativa. A análise foi mantida para fins ilustrativos, mas, em um cenário com dados reais e variabilidade representativa, a regressão linear poderia revelar tendências relevantes para a tomada de decisão.

Modelo de Celular que possui app PicMoney

O gráfico busca avaliar a relação entre o custo do modelo de celular dos usuários e a posse do aplicativo PicMoney. No eixo horizontal, tem-se a variável binária “Possui App PicMoney” (0 = não possui, 1 = possui), enquanto no eixo vertical está representado o custo do celular.



Foi ajustada uma reta de regressão linear, cuja equação é $y = 2.59 \times 10^3 - 7.87x$, com coeficiente de correlação $R = -0.003$, coeficiente de determinação $R^2 \approx 0$ e valor de significância $p = 0.418$.

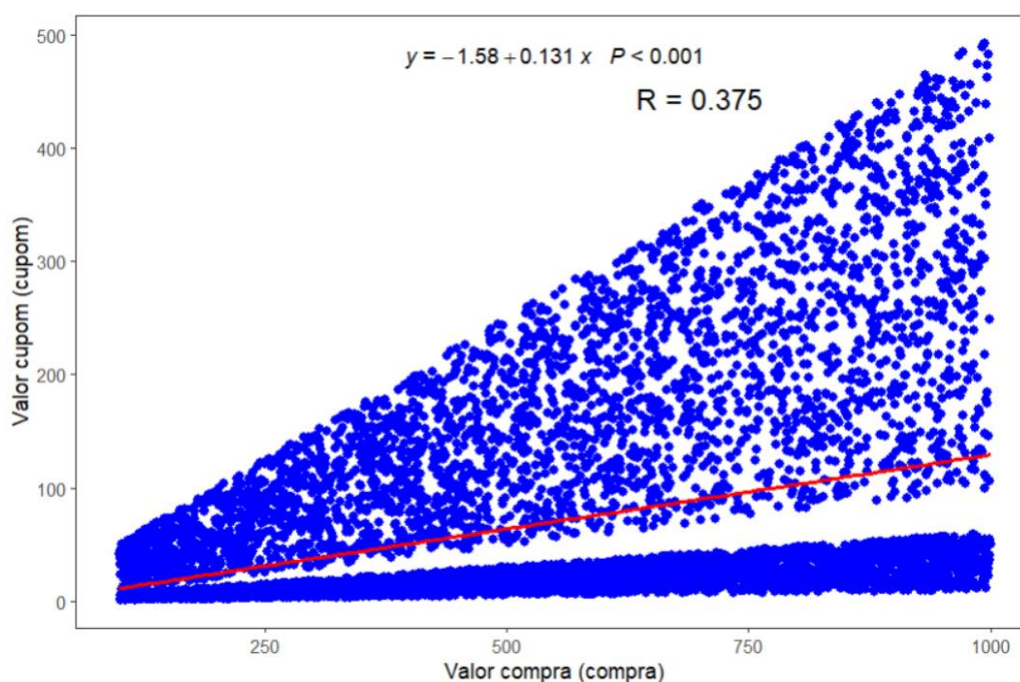
A reta ajustada apresentada confirma que os dados não possuem relação entre si. Mesmo que se pudesse supor que usuários com celulares mais caros tenderiam a adotar o aplicativo com maior ou menor frequência, o gráfico demonstra que a inteligência artificial que a PicMoney utilizou, responsável pela geração da base simulada, definiu a mesma probabilidade de ocorrência para todos os grupos, independentemente de características particulares como o valor do aparelho. Essa construção explica por que não há diferença perceptível entre os grupos e reforça que o coeficiente de correlação de Pearson próximo de zero indica ausência de associação entre as variáveis.

O coeficiente de correlação próximo de zero ($R = -0.003$) e o valor praticamente nulo de R^2 confirmam que a regressão não explica a variação dos dados. Além disso, o valor de $p = 0.418$ reforça a ausência de significância estatística.

Massa de Teste com Lojas e Valores

Nesta parte da análise, buscou-se identificar se existe relação entre o valor das compras realizadas e o valor dos cupons concedidos. O objetivo foi compreender se, à medida que o valor da compra aumenta, o valor do cupom também tende a crescer de forma proporcional. Para isso, foi aplicado um modelo de regressão linear, de modo a avaliar tanto a tendência média representada pela reta ajustada quanto a intensidade da correlação entre as variáveis.

Para a construção dos gráficos, foram utilizados dados coletados a partir de registros de compras e dos respectivos valores de cupons associados a cada transação. O procedimento consistiu em organizar essas informações em uma base de dados e, posteriormente, aplicar análises estatísticas que permitiram identificar a relação entre as variáveis envolvidas. Dessa forma, os gráficos apresentados refletem a distribuição dos dados reais e as tendências identificadas a partir deles.



Observação:

Ao analisarmos o estado do gráfico acima, observamos que há uma faixa em branco no meio do gráfico. Quando nós procuramos nos arquivos que a PicMoney disponibilizou uma informação que explicasse bem o porquê desse fenômeno, nós identificamos a seguinte resposta da IA que a empresa utilizou para gerar as bases de informações:

- **nome da loja e endereço:** escolhidos entre estabelecimentos reais do espigão da Av. Paulista e imediações.
- **valor da compra:** entre R\$ 100,00 e R\$ 1.000,00.
- **valor do cupom:** calculado conforme o tipo:
 - Produto → entre 1% e 6%
 - Desconto → entre 10% e 50%
 - Cashback → fixo de 5%

Observando o trecho acima, é possível concluir que os valores de cupom gerados pela IA são de Produto (entre 1% e 6%), Cashback (fixo de 5%) e Desconto (entre 10% e 50%). A partir disso depreende-se que a faixa entre 6% e 10% não possui valores, ou seja, no gráfico esse espaço fica em branco, formando aquela divisão entre tapete (piso de Cashback e Produto) e ladeira (Categoria de Desconto).

1. A linha vermelha

A linha vermelha foi desenhada usando `geom_smooth(method = "lm")`, que ajusta um modelo de regressão linear simples:

O resultado foi: $y = -1.58 + 0.131x$

Cálculo da Reta de Regressão Linear:

$$\begin{aligned}\hat{y} &= a + bx \\ b &= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \\ b &= \frac{460.741,43}{3.510.472,56} \approx 0,131 \\ a &= \bar{y} - b\bar{x} \\ a &= 80,94 - (0,131 \cdot 552,72) \approx -1,58 \\ \hat{y} &= -1,58 + 0,131x\end{aligned}$$

Interpretação:

- **Intercepto (-1.58):** quando valor_compra é zero, o modelo prevê um cupom próximo de -1.58 (valor que matematicamente sai da reta, mesmo que não faça sentido prático — não dá para ter cupom negativo, mas é normal em regressão).
- **Inclinação (0.131):** para cada R\$ 1,00 a mais de compra, o valor do cupom cresce em média R\$ 0,131.

2. O p-valor ($P < 0.001$)

O p-valor testa se a inclinação da reta é estatisticamente diferente de zero. Aqui $P < 0.001$, ou seja, a probabilidade da inclinação ser zero por acaso é muito baixa → existe relação significativa entre compra e cupom.

3. O R (coeficiente de correlação de Pearson)

Cálculo do R

$$\begin{aligned}
 R &= \frac{n \cdot \sum xy - \sum x \cdot \sum y}{\sqrt{n \cdot \sum x^2 - (\sum x)^2} \cdot \sqrt{n \cdot \sum y^2 - (\sum y)^2}} \\
 R &= \frac{10.000 \cdot 475.691.880,76 - 5.496.840,75 \cdot 703.536,66}{\sqrt{10.000 \cdot 3.702.426.048,92 - (5.496.840,75)^2} \cdot \sqrt{10.000 \cdot 132.416.739,02 - (703.536,66)^2}} \\
 R &= \frac{889.689.825.744,12}{2.374.396.105.369,72} \approx 0,3747 \\
 R^2 &= (R)^2 \\
 R &= 0,3747 \\
 R^2 &= (0,3747)^2 = 0,1404
 \end{aligned}$$

O resultado já está dado:

R=0.375

Ou seja, existe correlação positiva moderada entre Valor da Compra e Valor do Cupom.

- Como $R > 0$, a relação é positiva → quanto maior o valor da compra, maior tende a ser o valor do cupom.

Como $R = 0.375$ não está próximo de 1, significa que a correlação não é muito forte, mas é estatisticamente significativa (o $p < 0.001$ confirma isso).

- Ele mede o quanto os pontos seguem a tendência da linha reta.

- Valores próximos de 1 indicam uma relação linear forte; valores próximos de 0 indicam relação fraca.

O **R^2 mede quanto da variação de Y** (valor do cupom) pode ser explicada pela variação de X (valor da compra) usando o modelo de regressão linear.

- Ele é simplesmente o **quadrado do coeficiente de correlação de Pearson (R)**:

$$R^2 = R \times R$$

- O valor vai de **0 a 1** (ou de **0% a 100%** se multiplicado por 100).
- Um **$R^2 = 0.14$** significa que **14% da variação no valor do cupom é explicada pelo valor da compra.**

Cálculo:

$$\begin{aligned} R^2 &= (R)^2 \\ R &= 0,3747 \\ R^2 &= (0,3747)^2 = 0,1404 \end{aligned}$$

Ou seja: há muita dispersão, outros fatores além do valor da compra influenciam o valor do cupom.