

Projeto 1: Prevendo Demanda de um Catálogo

Passo 1: Compreensão do Negócio e dos Dados

Decisões Chaves:

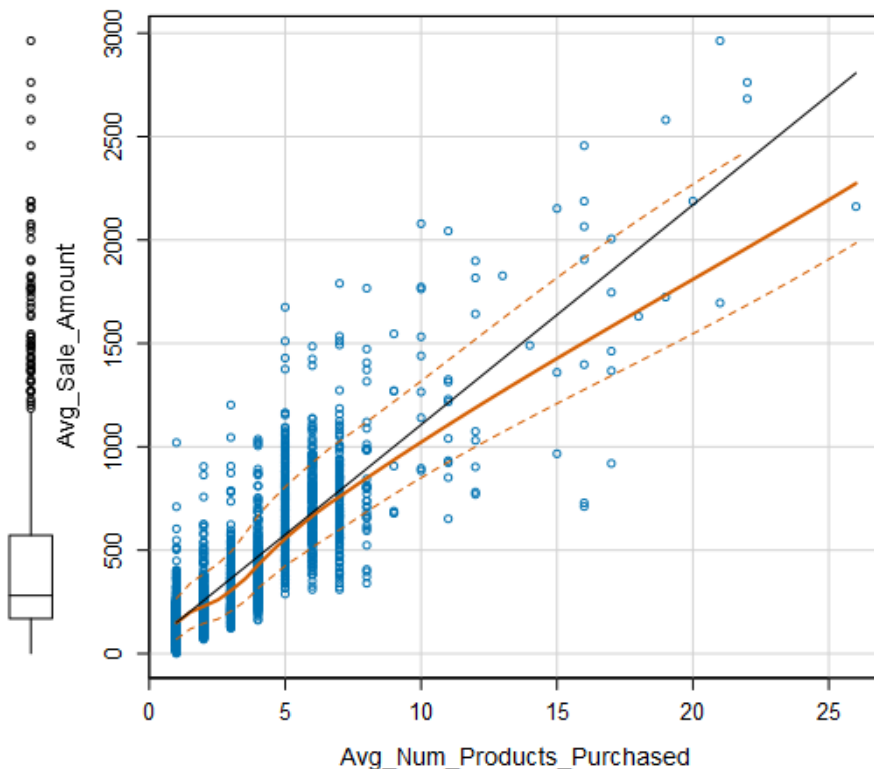
É preciso decidir se é vantajoso enviar o catálogo para os novos clientes, baseado nos custos e no possível lucro da operação.

É necessário estimar a receita geradas a partir do envio do catálogo, para comparar essa informação com o seu custo de impressão e distribuição.

Passo 2: Análise, modelagem e validação

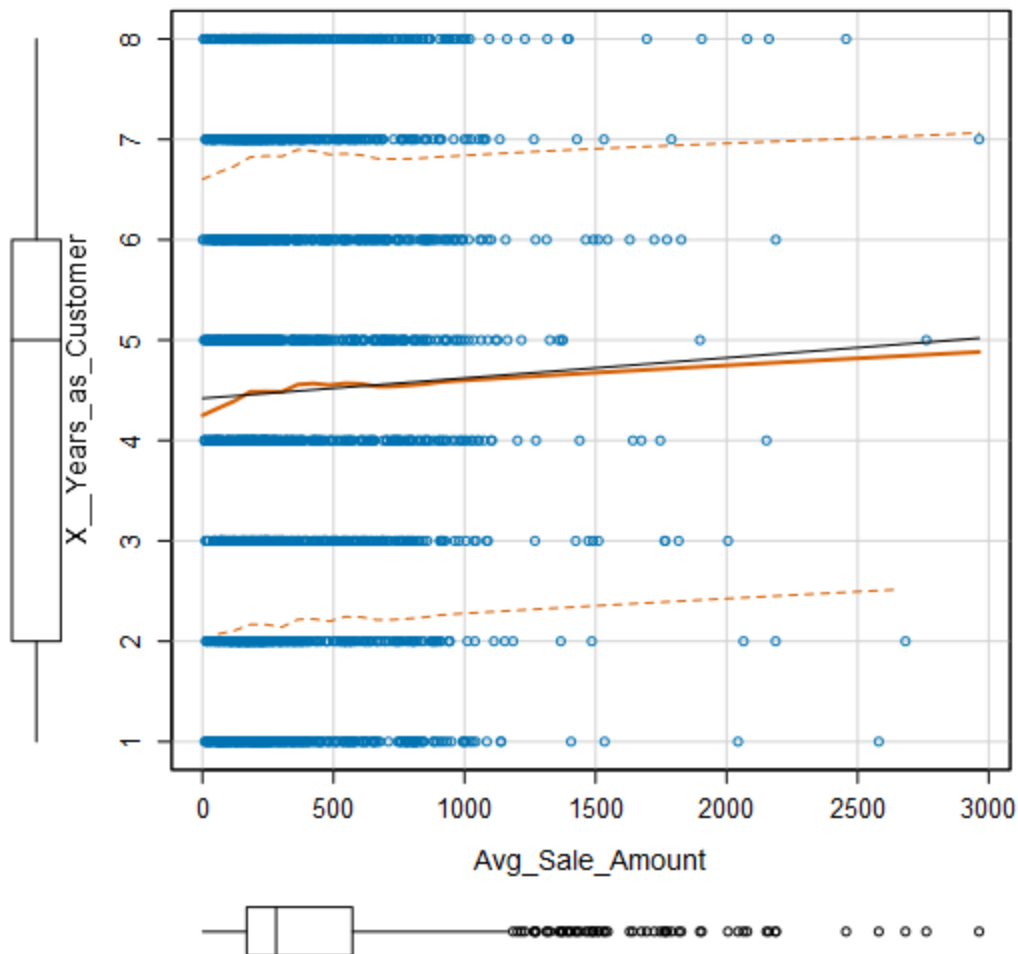
As variáveis de previsão escolhidas foram “Customer Segment” e “Avg Num Products Purchased”, pois essas foram as variáveis que apresentaram p-valor menor que 0,05 e significância alta na regressão linear, com “Avg Sale Amount” sendo a variável-alvo. É possível confirmar essa correlação forte plotando o gráfico de dispersão entre as variáveis contínuas “Avg Num Products Purchased” e “Avg Sale Amount”.

A variável “Responded to last Catalog” também tem uma relação forte na regressão linear, mas como essa informação não se aplica à lista nova ela foi descartada.



Se incluirmos a variável “Years as Customers” no modelo, podemos verificar que o seu valor é próximo de 0,05, mas há pouca significância. Se plotarmos o gráfico de dispersão entre as variáveis, fica claro que a relação entre elas é bem mais fraca se comparada ao exemplo anterior.

Scatterplot of Avg_Sale_Amount versus X_Years_as_Customer



Os p-valores de todas as variáveis escolhidas, portanto, são menores que 0,05 e com alta significância, o que indica um alto nível de confiabilidade:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	303.46	10.576	28.69	< 2.2e-16	***
Customer.SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16	***
Customer.SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16	***
Customer.SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16	***
Avg.Num.Products.Purchased	66.98	1.515	44.21	< 2.2e-16	***

Os valores R-Quadrado e R-Quadrado Ajustado são 0,8369 e 0,8366, respectivamente,

o que significa que o modelo é bom.

A melhor equação de regressão linear com base nos dados disponíveis é:

$$\text{ASA} = 303,46 - 149,36 * \text{LCO} + 281,84 * \text{LC\&CC} + 0 * \text{CCO} - 245,42 * \text{SML} + 66,98 * \text{ANPP}$$

Onde:

ASA: Average Sale Amount

LCO: Customer.SegmentLoyalty Club Only

LC&CC: Customer.SegmentLoyalty Club and Credit Card

CCO: Customer.SegmentCredit Card Only

SML: Customer.SegmentStore Mailing List

ANPP: Avg.Num.Products.Purchased

Passo 3: Apresentação/Visualização

Use os resultados do modelo para fornecer uma recomendação. (limite de 500 palavras)

A empresa deve enviar o catálogo para os clientes, uma vez que o retorno esperado desta ação é um lucro de US\$ 21987,44, um valor maior do que o limite inferior mínimo definido pela gerência.

Esse valor foi encontrado ao aplicar ao mailing list o modelo de previsão baseado nos clientes antigos, que aponta um valor médio que este cliente em potencial deve gastar. Foi levada em conta também a probabilidade do novo cliente efetuar a compra, para chegar a um valor total de quanto de receita se pode esperar do grupo.

Com esse valor em mãos, bastou deduzir a margem bruta e o custo de produção e envio dos catálogos, para se chegar ao valor final correspondente ao lucro.