

Projeto 1: Prevendo Demanda de um Catálogo

Passo 1: Compreensão do Negócio e dos Dados

Decisões Chaves:

1. Que decisões precisam ser feitas??

A principal decisão a ser feita no projeto é enviar ou não o catálogo de produto para 250 clientes sendo que o lucro deve ser superior a US\$ 10.000. Para isso, devemos criar um modelo preditivo que, com base no histórico da empresa, prevê a receita total desses 250 clientes que receberão o catálogo.

2. Que dados são necessários para subsidiar essas decisões??

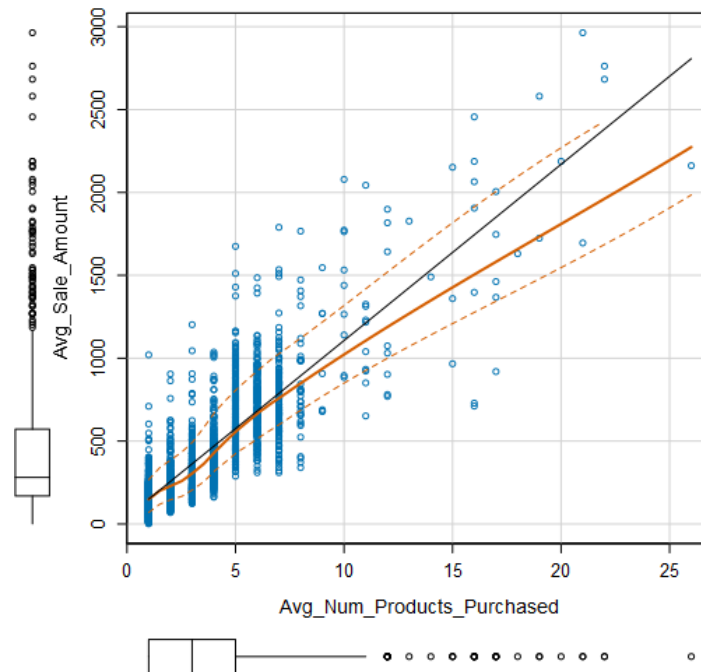
São necessárias para subsidiar essas decisões as informações de vendas passadas através dos catálogos. Nessa base de dados, podemos ver a forma de pagamento, o número médio de produtos comprados, o cliente, endereço, e outras variáveis. Além disso, precisamos do preço do catálogo, da margem de lucro e da probabilidade do cliente comprar pelo catálogo. E por fim, precisamos aplicar o modelo na base de dados que contém os 250 clientes.

Passo 2: Análise, modelagem e validação

1. Como e por que você selecionou as variáveis de previsão (veja texto suplementar) em seu modelo? Você deve explicar como as variáveis de previsão contínuas que você escolheu têm uma relação linear com a variável-alvo. Você deve incluir gráficos de dispersão em sua resposta.

As variáveis foram escolhidas através de análises feitas por gráficos de dispersão e pelos valores de p-valor das variáveis. O gráfico de dispersão que melhor se adequou ao modelo foi o número médio de produtos comprados em função das vendas.

Plot of Avg_Num_Products_Purchased versus Avg_Sale



A segunda variável que pode ser considerada forte no modelo foi a Segmentação dos clientes. Ela aumentou o valor do R quadrado e possui ótimo p-valor.

2. Explique por que você acredita que seu modelo linear é um bom modelo. Você deve justificar o seu raciocínio usando os resultados estatísticos criados pelo seu modelo de regressão. Para cada variável selecionada, por favor justificar por que cada variável é uma boa opção para o seu modelo, usando os valores-p e valores R-quadrado produzidos pelo seu modelo.

O modelo aplicador é um bom modelo por apresentar um R quadrado ajustado igual a 0.8366. Além disso, os valores-p são inferiores a 0.05 conforme a figura:

	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	***

3. Qual é a melhor equação de regressão linear com base nos dados disponíveis? Cada coeficiente não deve ter mais de 2 dígitos após o decimal (ex: 1,28)

$$Y = 303.46 - 149.36 * \text{Customer_SegmentLoyalty Club Only} + 281.84 * \text{Customer_SegmentLoyalty Club and Credit Card} - 245.42 * \text{Customer_SegmentStore Mailing List} + 66.98 * \text{Avg_Num_Products_Purchased}$$

Passo 3: Apresentação/Visualização

1. Qual é a sua recomendação? A empresa deve enviar o catálogo para estes 250 clientes?

Sim, a empresa deve enviar o catálogo para os 250 clientes pela estimativa de um lucro superior a US\$ 10,000.

2. Como você chegou na sua recomendação? (Por favor, explique a sua lógica para os revisores poderem lhe dar feedback sobre o seu processo)

Após calcular a receita das vendas, multipliquei a receita pela probabilidade de a pessoa comprar o catalogo também. Depois tirei 50% da margem bruta. A última operação foi subtrair o preço do catálogo de US\$6.50. Dessa forma é possível determinar o lucro total da empresa para os 250 clientes.

3. Qual é o lucro esperado do novo catálogo (assumindo que o catálogo é enviado para estes 250 clientes)?

O lucro do novo catálogo é igual a US\$ 21,987.43