

Faça o download para poder editar

Projeto 1: Prevendo Demanda de um Catálogo

Complete cada seção. Quando estiver pronto, salve o arquivo como um documento PDF e envie-o aqui: <https://classroom.udacity.com/nanodegrees/nd008/parts/c0b53068-1239-4f01-82bf-24886872f48e/project>

Passo 1: Compreensão do Negócio e dos Dados

Fornecer uma explicação das decisões importantes que precisam ser feitas. (limite de 500 palavras)

Decisões Chaves:

Responda estas perguntas

1. Que decisões precisam ser feitas??

R: Saber se vale a pena enviar o catálogo para os 250 clientes. Saber se vale a pena = Se dará lucro.

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

R: Dados do último catálogo, quem comprou, média de compra, anos como cliente.

Passo 2: Análise, modelagem e validação

Forneça uma descrição de como você configurou o seu modelo de regressão linear, quais as variáveis usadas e o porquê, assim como os resultados do modelo. Visualizações são incentivadas. (limite de 500 palavras)

Importante: Use o `p1-customers.xlsx` para treinar o modelo linear.

No mínimo, responda à estas perguntas:

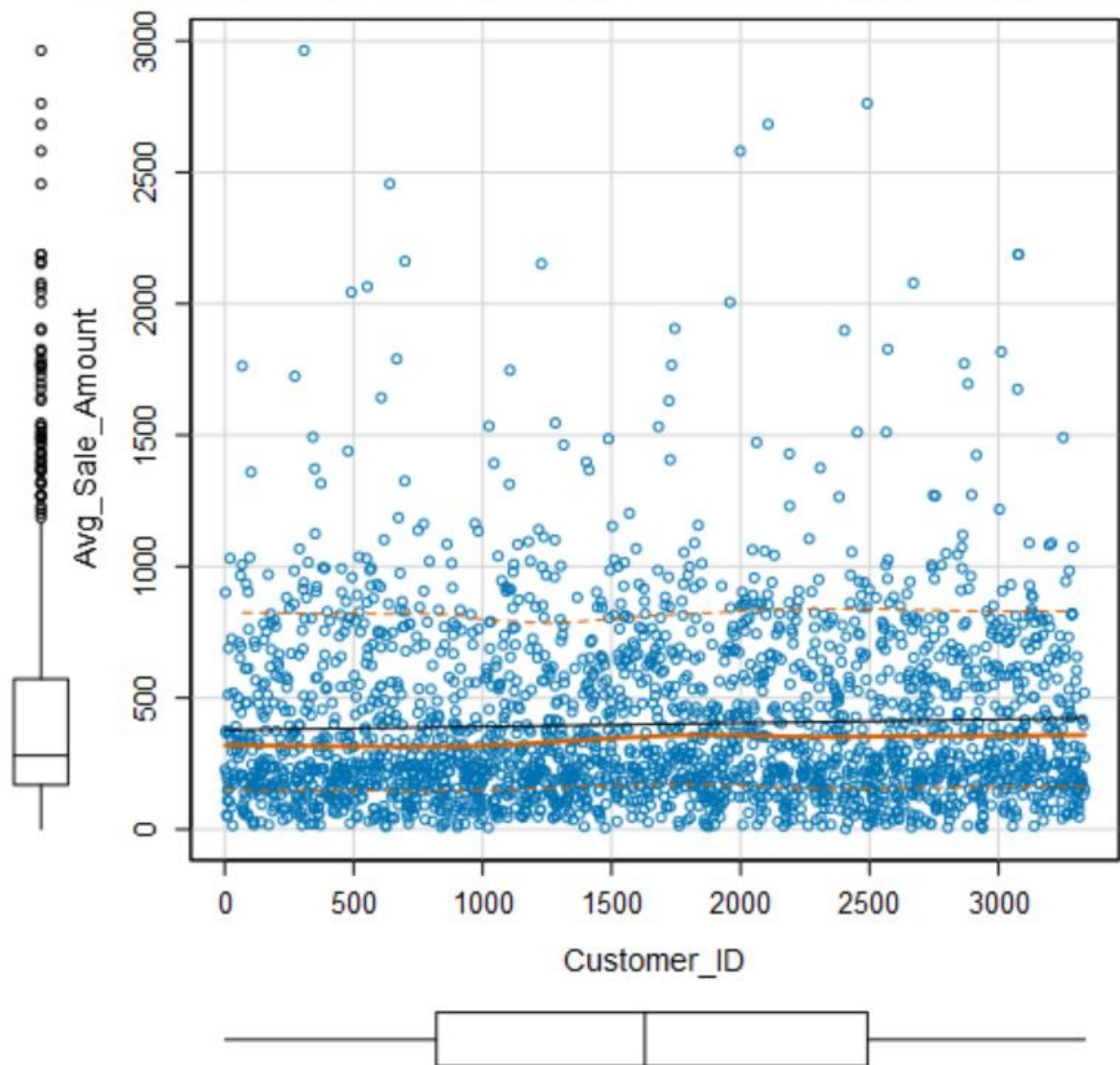
1. Como e por que você selecionou [as variáveis de previsão \(veja texto suplementar\)](#) em seu modelo? Você deve explicar como as variáveis de previsão contínuas que você escolheu têm uma relação linear com a variável-alvo. Consulte esta [lição](#) para ajudar você a explorar seus dados e usar gráficos de dispersão para procurar relações lineares. Você deve incluir gráficos de dispersão em sua resposta.

R: A primeira ideia foi utilizar todas as variáveis e ver como elas se ligam ao modelo. Analisei o P Valor e retirei as variáveis com p valor acima de 0,05.

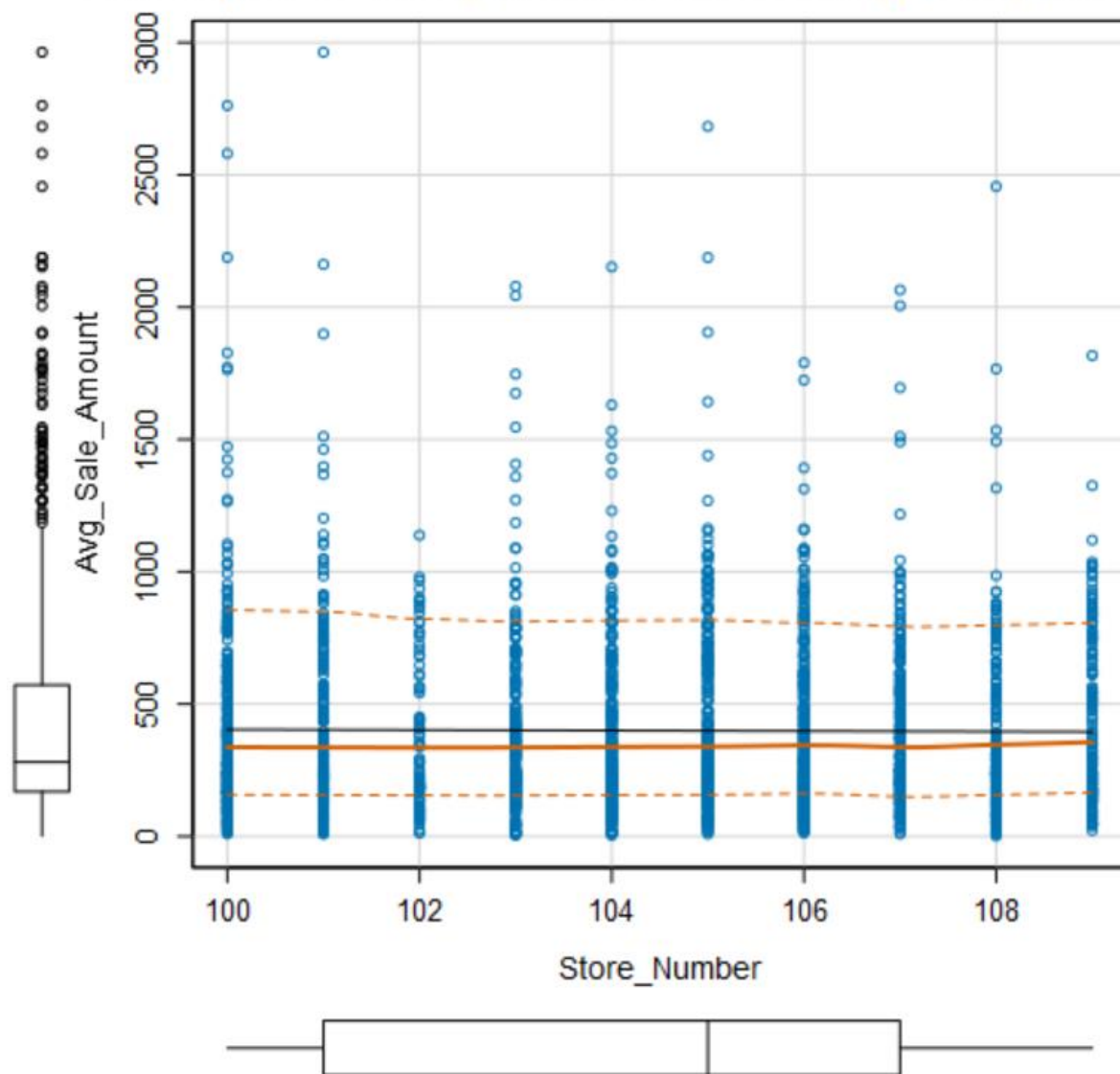
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.176e+02	12.764602	24.884	< 2.2e-16	***
Customer.SegmentLoyalty Club Only	-1.499e+02	8.966376	-16.716	< 2.2e-16	***
Customer.SegmentLoyalty Club and Credit Card	2.821e+02	11.918073	23.671	< 2.2e-16	***
Customer.SegmentStore Mailing List	-2.429e+02	9.811708	-24.757	< 2.2e-16	***
Customer.ID	-1.543e-03	0.002938	-0.525	0.59961	
Responded.to.Last.CatalogYes	-2.818e+01	11.261656	-2.502	0.01242	*
Avg.Num.Products.Purchased	6.689e+01	1.516117	44.116	< 2.2e-16	***
X..Years.as.Customer	-2.303e+00	1.221942	-1.885	0.05956	.

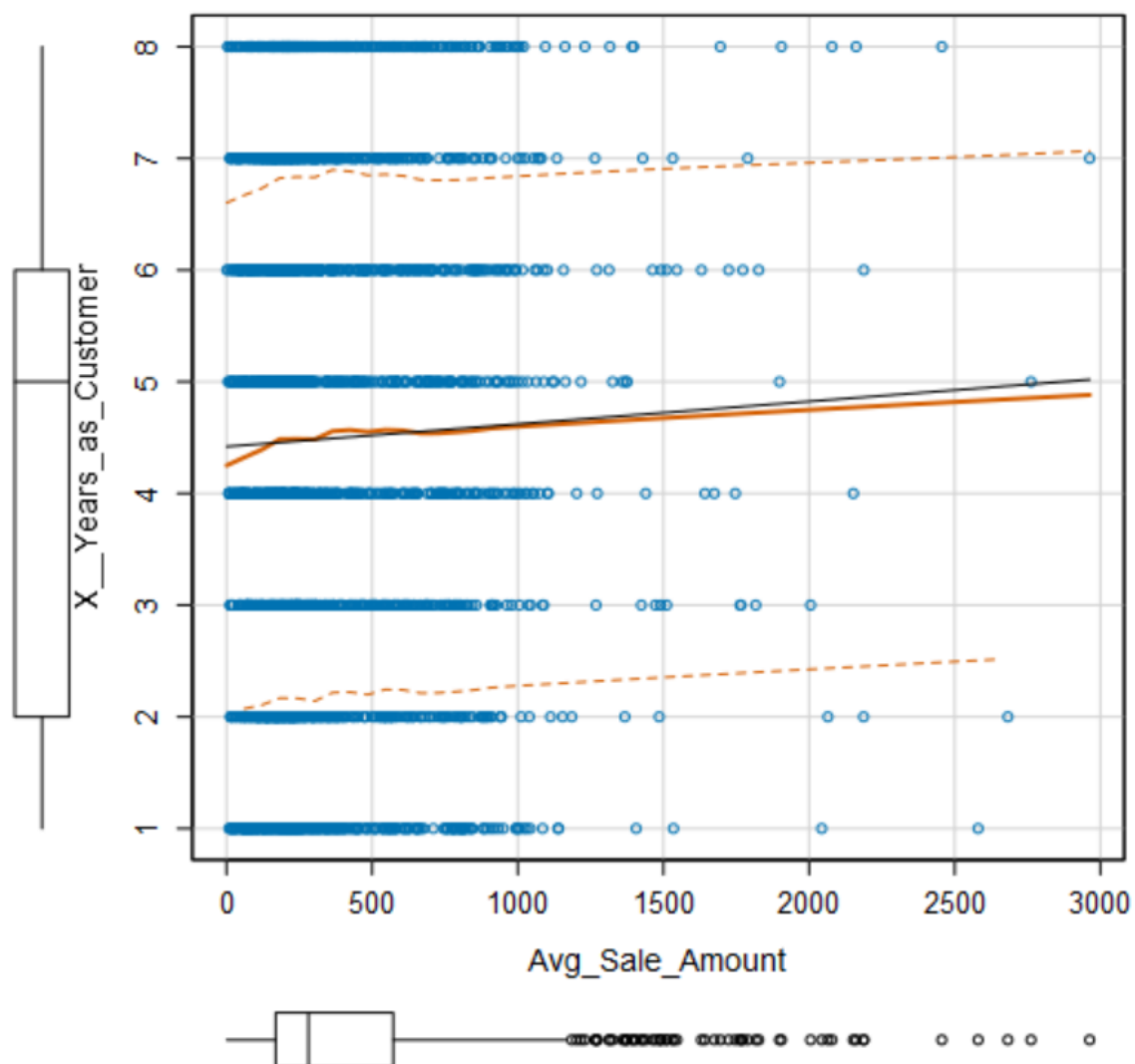
Scatterplot of Customer_ID versus Avg_Sale_Amount



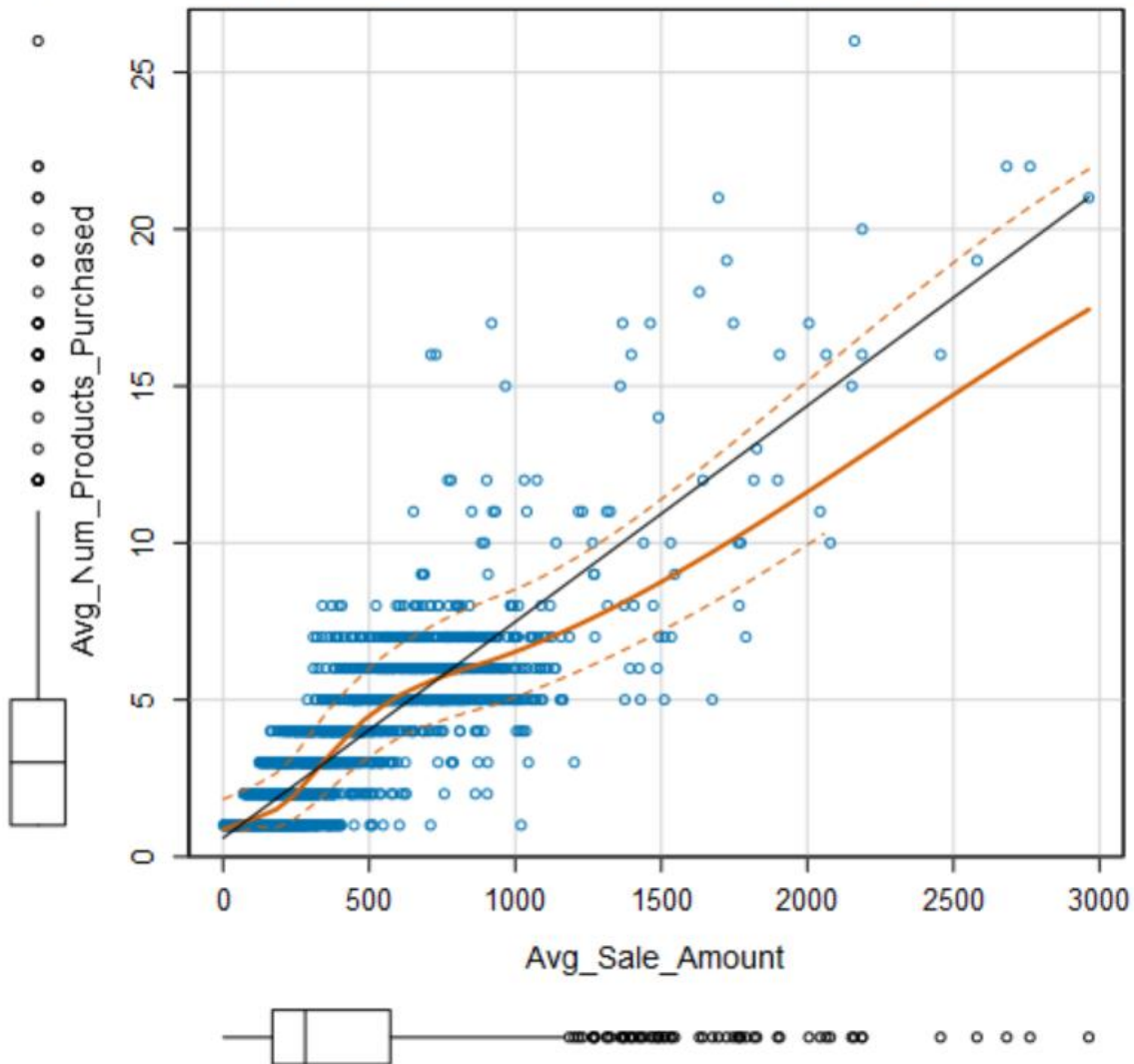
Scatterplot of Store_Number versus Avg_Sale_Amount



Scatterplot of Avg_Sale_Amount versus X_Years_as_Customer



Scatterplot of Avg_Sale_Amount versus Avg_Num_Products_P



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.

R: Todas as variáveis utilizadas no modelo possuem p valor abaixo de 0,05 e o R-Quadrado é próximo de 1 e acima de 0,7, validando o modelo linear como um bom modelo.

Residual standard error: 137.48 on 2370 degrees of freedom
 Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366
 F-statistic: 3040 on 4 and 2370 degrees of freedom (DF), p-value < 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)

Importante: A equação de regressão deve estar na forma:

$$Y = \text{Intercept} + b1 * \text{Variable_1} + b2 * \text{Variable_2} + b3 * \text{Variable_3} \dots$$

Por exemplo: $Y = 482.24 + 28.83 * \text{Loan_Status} - 159 * \text{Income} + 49 (\text{Se Type: Credit Card}) - 90 (\text{Se Type: Mortgage}) + 0 (\text{Se Type: Cash})$

Note que **devemos** incluir o coeficiente 0 para o type Cash.

$$\begin{aligned} R: & 303.36 + (-149.36) * (\text{Customer.SegmentLoyalty Club Only}) + 281.84 * \\ & (\text{Customer.SegmentLoyalty Club and Credit Card}) + (-245.42) * \\ & (\text{Customer.SegmentStore Mailing List}) + 66.98 (\text{Avg.Num.Products.Purchased}) \end{aligned}$$

Nota: Para os alunos que utilizam outro software que não Alteryx, se você decidir usar Customer Segment como uma das suas variáveis de previsão, por favor, defina o caso base apenas para Credit Card.

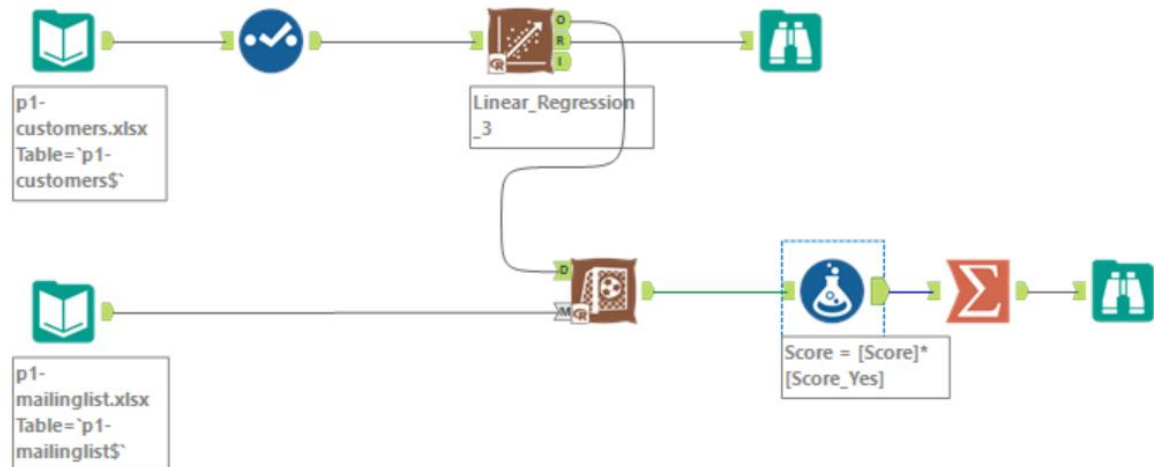
Passo 3: Apresentação/Visualização

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

No mínimo, responder à estas perguntas:

1. Qual é a sua recomendação? A empresa deve enviar o catálogo para estes 250 clientes?
R: Sim, a empresa terá lucro com o envio dos catálogos.
2. Como você chegou na sua recomendação? (Por favor, explique a sua lógica para os revisores poderem lhe dar feedback sobre o seu processo)

R: Ao utilizar o modelo de regressão linear, anexe a base de dados mailinglist a ferramenta score para utilizar a fórmula da regressão na outra base de dados. Após isso, utilizei a ferramenta fórmula para calcular o valor a probabilidade de compra dos clientes, chegando ao resultado da receita de \$ 47,224.87.



A receita esperada com a distribuição do catálogo para os 250 clientes é de \$ 47,224.87. O custo é de 50% a receita + 6,50 por catálogo. Ou seja, \$23,612.43 + 1625 = 25,237.43

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

R: A receita esperada com a distribuição do catálogo para os 250 clientes é de \$ 47,224.87. O custo é de 50% a receita + 6,50 por catálogo. Ou seja, \$25,237.43 de custo, tendo como lucro \$ 21,987.44.

Antes de enviar

Por favor verifique se suas respostas estão de acordo com o que é pedido na [rubrica](#). Os revisores vão usar esta rubrica para avaliar o seu trabalho.