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. Através de análises de variáveis, se o lucro for maior que $ 10,000.00, valerá a pena enviar o catálogo.

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

R: Dados do último catálogo, quem comprou, média de compra, anos como cliente. É importante entender dados de compras anteriores, quem comprou, número de itens comprado, dinheiro gasto por cliente.

Além desses dados, para poder dar a resposta final, é preciso dos dados como valor do catálogo, custo por catálogo e probabilidade de compra.  
  
Poderia ser utilizado também dados geográficos.

## Passo 2: Análise, modelagem e validação

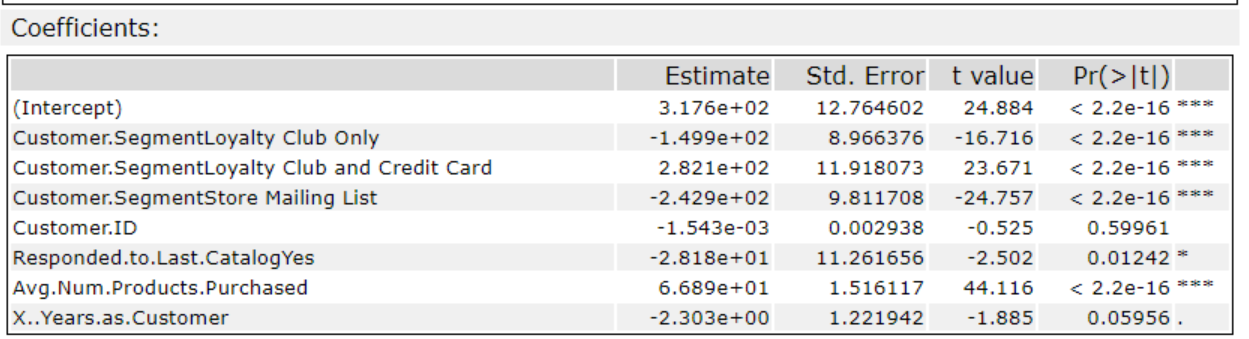
*Forneça uma descrição de como você configurou o seu modelo de regressão linear, quais as variáveis ​​usadas e o por quê, assim como os resultados do modelo. Visualizações são incentivadas. (limite de 500 palavras)*

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

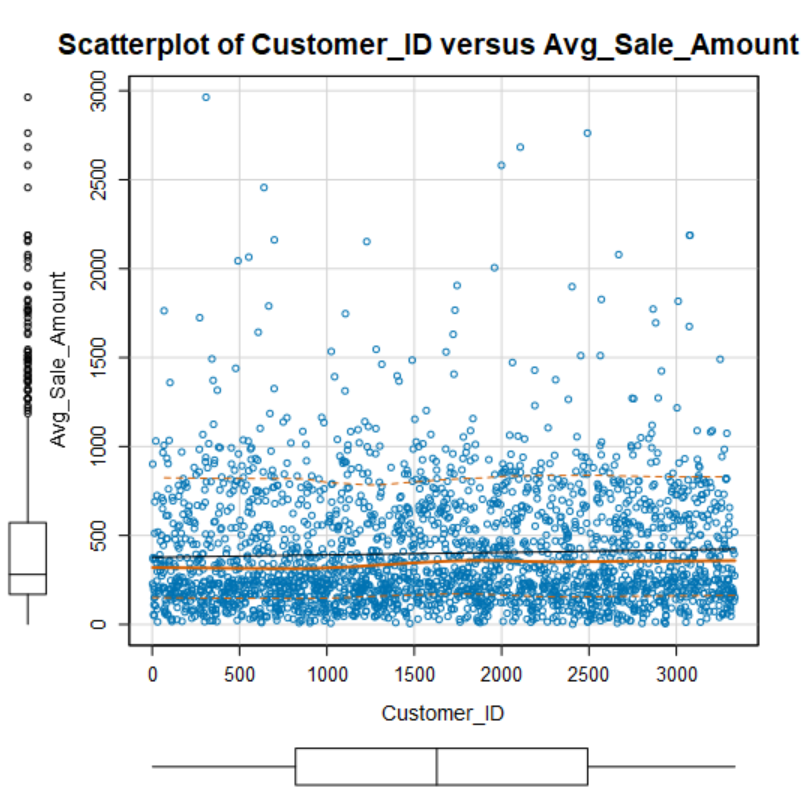
*No mínimo, responda à estas perguntas:*

1. Como e por que você selecionou as variáveis de previsão [(veja texto suplementar)](https://classroom.udacity.com/courses/ud976/lessons/4e33b70a-72a4-47cb-959a-28632ae6aaff/concepts/631d190c-8626-4dd7-92df-f5bd96913c48) 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](https://classroom.udacity.com/nanodegrees/nd008/parts/c0b53068-1239-4f01-82bf-24886872f48e/modules/bf705147-0d7c-4492-842a-698a6410a8a3/lessons/4e33b70a-72a4-47cb-959a-28632ae6aaff/concepts/631d190c-8626-4dd7-92df-f5bd96913c48) 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.

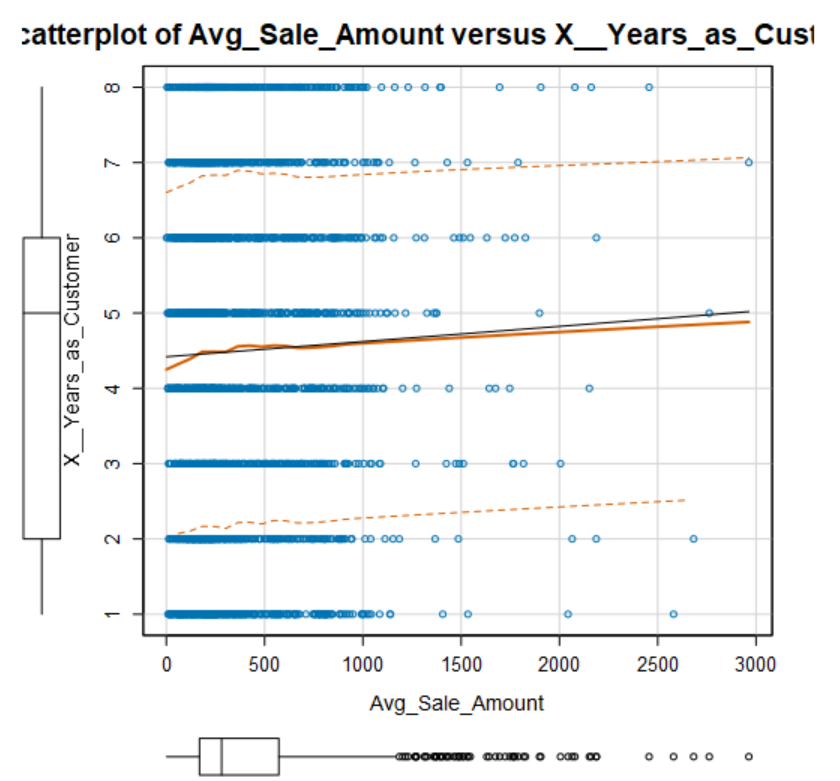
R: A primeira ideia foi utilizar todas as variáveis e ver como elas se ligam ao modelo. Analisei o P Valor e retirei as variáveis com p valor acima de 0,05.

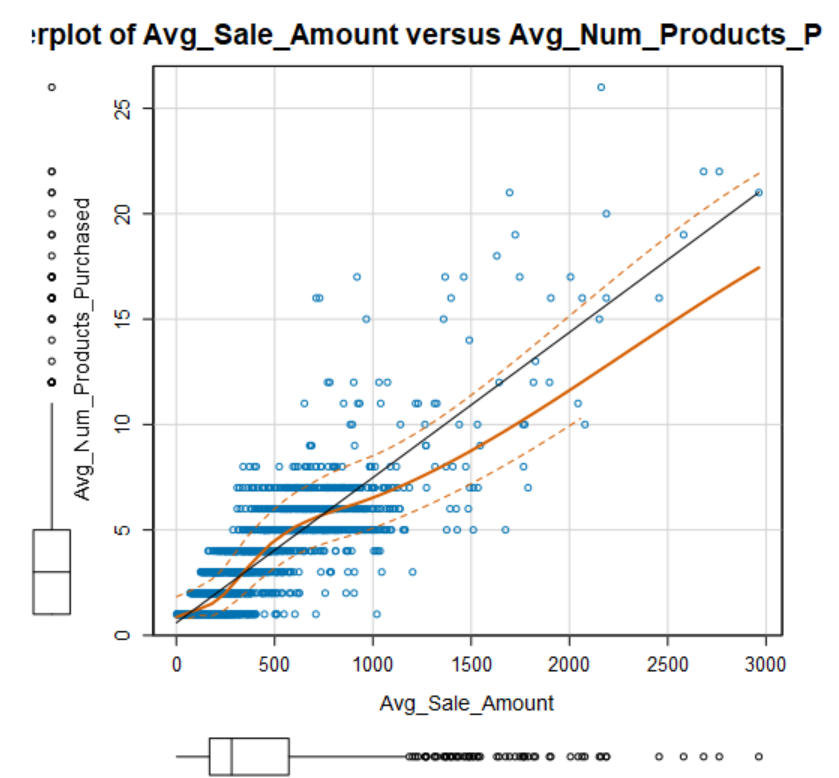


Conforme a imagem com os coeficientes, percebi que algumas variáveis não seriam boas para o modelo, sendo elas, customer.id, responded.to.last.catalogyes, years.as.customer.  
  
Por isso, utilizei as variáveis Customer Segment e Avg.Num.Products.Purchased.



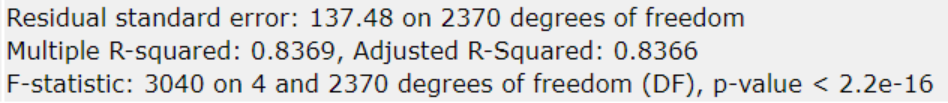






1. Explique por que você acredita que seu modelo linear é um bom modelo. Você deve justificar o seu raciocínio usando os resultados estatísticos criados pelo seu modelo de regressão. Para cada variável selecionada, por favor justificar por que cada variável é uma boa opção para o seu modelo, usando os 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.



1. 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 = Intercept + b1 \* Variable\_1 + b2 \* Variable\_2 + b3 \* Variable\_3……*

**Por exemplo:** Y = 482.24 + 28.83 \* Loan\_Status – 159 \* Income + 49 (Se Type: Credit Card) – 90 (Se Type: Mortgage) + 0 (Se Type: Cash)

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

R: *303.36 + (-149.36) \* (Customer.SegmentLoyalty Club Only) + 281.84 \* (Customer.SegmentLoyalty Club and Credit Card) + (-245.42) \* (Customer.SegmentStore Mailing List) + 66.98 (Avg.Num.Products.Purchased)*

**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

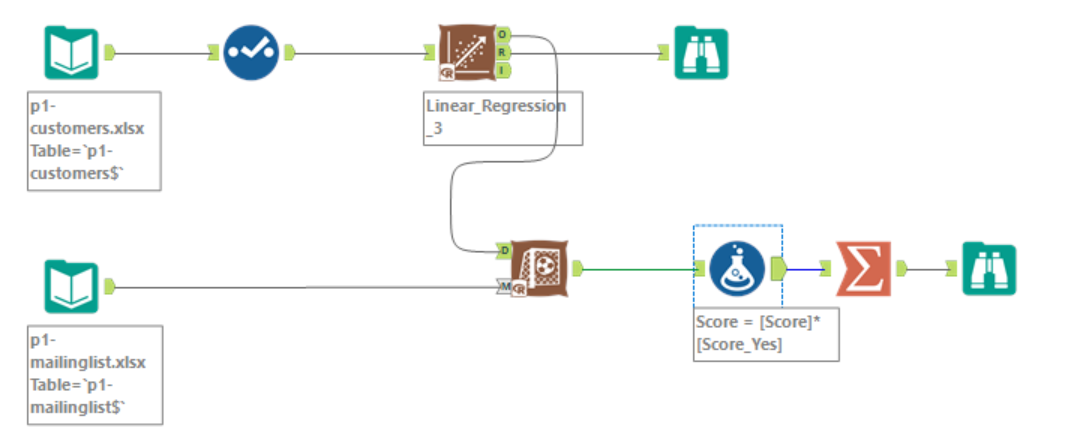
*Use os resultados do modelo para fornecer uma recomendação. (limite de 500 palavras)*

*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.

1. 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 linar, anexei 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

1. 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](https://review.udacity.com/#!/rubrics/639/view). Os revisores vão usar esta rubrica para avaliar o seu trabalho.