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??

Precisamos saber se seguir com a opção de enviar o catálogo para os novos clientes reverterá no lucro esperado pela gerência da empresa. Neste caso, a gerência espera obter um lucro de no mínimo US$10.000 para que seja viável enviar os catálogos ou não. Não podemos esquecer de considerar o custo de produção dos produtos e o custo de impressão e distribuição por catálogo.

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

Neste caso, precisamos de dados históricos da performance dos nossos clientes, devemos ter algumas variáveis essenciais para a construção do modelo, por exemplo o Valor Médio de Vendas, Custo de Produção dos Catálogos, Segmento do Cliente e o Número Médio dos Produtos Comprados e , pois a partir dessas variáveis treinaremos o modelo para usá-lo na base de dados que será feita a previsão.

## Passo 2: Análise, modelagem e validação

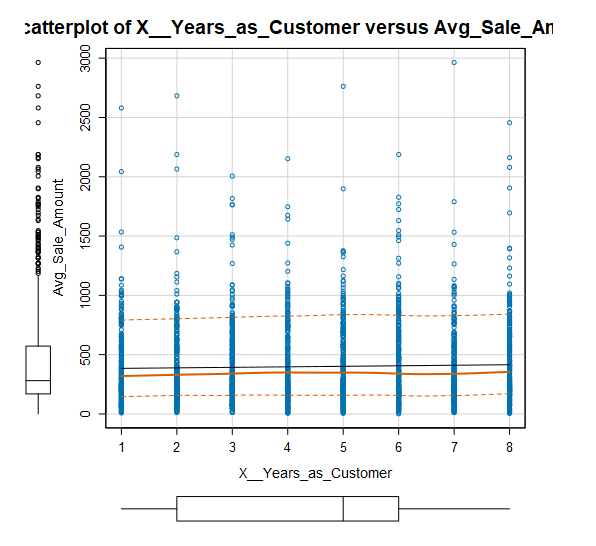
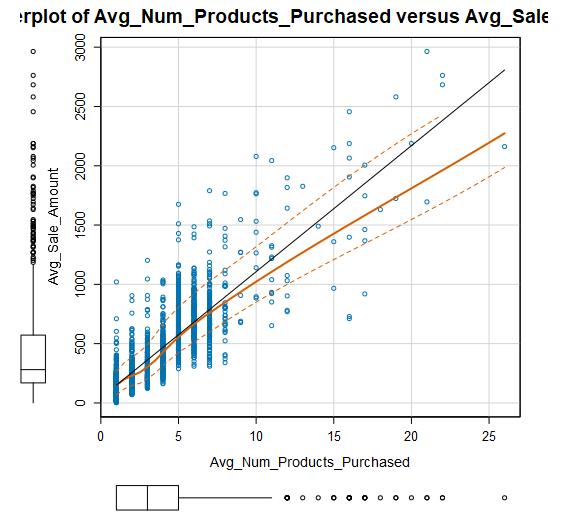
*Forneça uma descrição de como você configurou o seu modelo de regressão linear, quais as variáveis ​​usadas e o por quê, assim como os resultados do modelo. Visualizações são incentivadas. (limite de 500 palavras)*

***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.

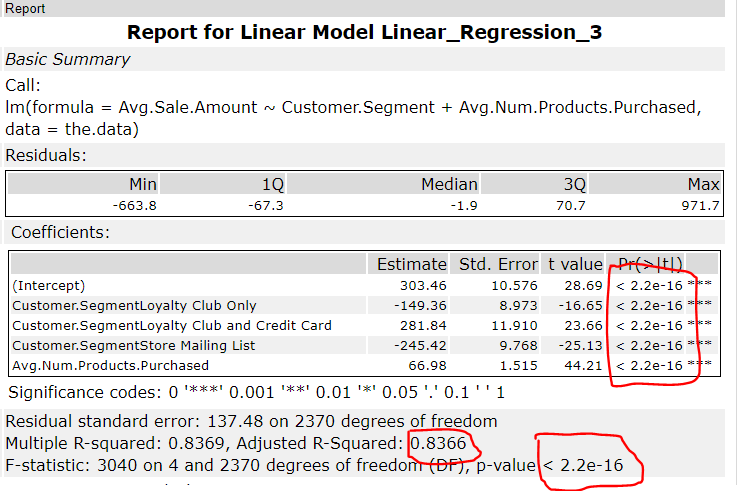
Escolhi a variável AVG\_NUM\_PRODUCTS\_PURCHASED para o modelo, pois vemos uma linearidade positiva no gráfico de dispersão, com isso podemos assumir que quanto maior o número em AVG\_NUM\_PRODUCTS\_PURCHASED, maior será nosso número em AVG\_SALE\_AMOUNT, variável está que queremos prever. Desconsiderei as outras variáveis contínuas pois não encontrei variáveis com força no modelo, como por exemplo a variável YEAS\_AS\_CUSTOMER, segundo gráfico. Como a linha neste gráfico é plana, podemos concluir que esta variável não é boa para a nossa modelagem. Além dessa variável, foi testado todas as variáveis categóricas e a única que foi considerada estatisticamente significativa para o modelo foi a variável CUSTOMER\_SEGMENT. A variável foi considera por obter um p-valor menor que 0.05.



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.

Depois de escolher as variáveis AVG\_NUM\_PRODUCTS\_PURCHASED e CUSTOMER\_SEGEMENT, podemos dizer que o modelo é bom, pois temos um resultado de p-valor inferior à 0.05 e isso mostra evidências que as variáveis escolhidas estão relacionadas com o valor que queremos prever, além disso todas variáveis também têm p-valor inferior à 0.05.

Além disso, temos um R² Ajustado muito bom, que explica 0.8366 da variabilidade da variável AVG\_SALE\_AMOUNT\_PREDICIT. Normalmente acima de 0.7 de R², podemos considerar que nossa variabilidade é explicada pelas nossas variáveis.



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)

*Y = 303.46 – 149.36 \* Loyalty\_Club\_Only + 281.84 \* Loyalty\_Club\_and\_Credit\_Card – 245.42 \* Store\_Mailing\_List + 66.98 \* Products\_Purchased + 0\*Credit\_Card\_Only*

**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.

**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?

Recomendo que o catálogo seja enviado para todos os 250 clientes, porém com algumas ressalvas. Como o lucro alvo era de US$ 10.000 e nesse modelo temos uma previsão de lucro de 120% maior que o alvo, ou seja, o nosso modelo previu um lucro de US$ 21.987,40.

Para resumir, recomendaria o envio do catálogo diante do lucro previsto pelo nosso modelo.

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

Após usar um modelo de regressão linear múltipla, usando as utilizando AVG\_NUM\_PRODUCTS\_PURCHASED e CUSTOMER SEGMENT, para calcular a média de vendas por cliente, multipliquei o valor que obtive pela fórmula de regressão com a variável SCORE\_YES (Probabilidade do cliente comprar), após isso multipliquei o novo valor por 0.50 (Pois temos um custo de produção de 50% para todos os produtos) e por fim subtrai $6.5 para cada linha de cliente (Preço de produção do catálogo).

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

Temos uma previsão de lucro de US$ 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.