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??

Uma vez que temos um problema de negócios onde se deseja realizar previsões de valores, devemos realizar análises preditivas e numéricas;

Os resultados dos dados são de natureza contínua, sendo assim, o modelo de regressão será utilizado para construir o modelo;

Após tendo o modelo sido montado e obtida a equação de regressão, esta será utilizada para predizer as vendas de cada pessoa na lista de e-mail;

A partir dos dados do passo anterior, será analisado se a empresa deve ou não enviar os catálogos para a lista de e-mail.

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

Dados de Customer Segment , Avg Sale Amount e Avg Num Products Purchased do arquivo p1-customers.xlsx;

Dados de Customer Segment e Avg Num Products Purchased do arquivo p1-mailinglist.xlsx;

Lucro esperado de cada catálogo (calculado).

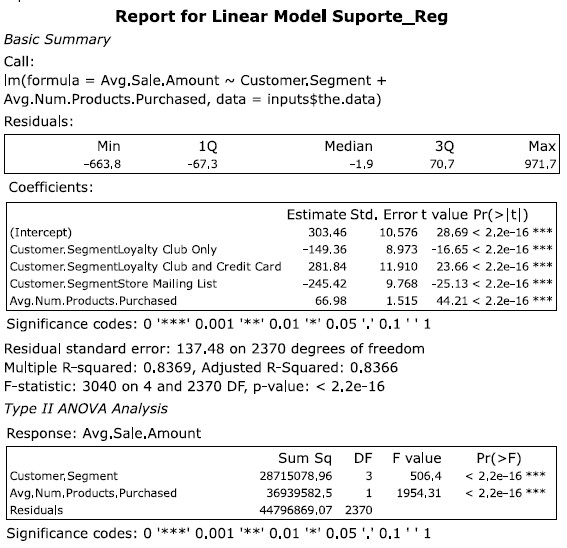
## Passo 2: Análise, modelagem e validação

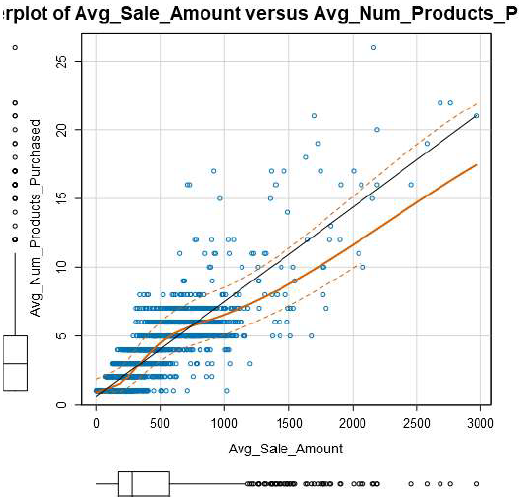
*Forneça uma descrição de como você configurou o seu modelo de regressão linear, quais as variáveis ​​usadas e o por quê, assim como os resultados do modelo. Visualizações são incentivadas. (limite de 500 palavras)*

***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.





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.

O valor Multiple R-squared de 0,8369 e o R-squared ajustado de 0,8366 são suficientemente altos e as variáveis preditivas são altamente significativas conforme os p-values.

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.

**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.

Y = 303.46 + (281.84 X Customer\_SegmentLoyalty Club and Credit Card) + (-149.36 X Customer\_SegmentLoyalty Club Card Only) + (-245.42 X Customer\_SegmentStore Mailing List) + ( 66.98 \* Avg\_Num\_Products\_Purchased) + Credit Card \* 0

Passo 3: Apresentação/Visualização

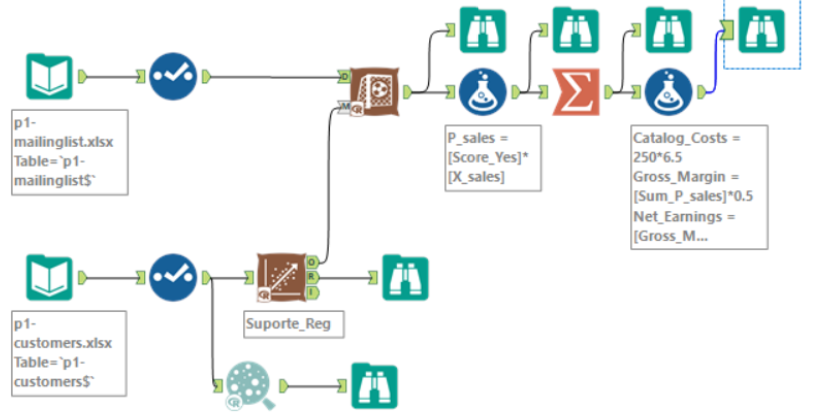
*Use os resultados do modelo para fornecer uma recomendação. (limite de 500 palavras)*

*No mínimo, responder à estas perguntas:*

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

Sim, pois o lucro previsto é maior que US $ 10.000.

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



O modelo foi desenvolvido com base no software Alteryx, conforme o diagrama de fluxo da imagem acima;

A ferramenta de Score gerou a predição das vendas para os 250 clientes presentes na lista de e-mails;

Em seguida, a predição das vendas foi multiplicada pela variável score\_yes correspondente;

Foi somado todas as vendas e então multiplicada pela margem bruta de 50%;

Por fim, foi subtraído da soma anterior o valor dos catálogos (250\*$6.50);

Assim, o lucro é superior a US $ 10000 e, portanto, rentável para enviar catálogo para os novos clientes.

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



O lucro previsto é 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.