

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 prever 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.

2. 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 \(veja texto suplementar\)](#) 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.

## **Report for Linear Model Suporte\_Reg**

### *Basic Summary*

Call:

```
lm(formula = Avg.Sale.Amount ~ Customer.Segment +
  Avg.Num.Products.Purchased, data = inputs$the.data)
```

Residuals:

Min	1Q	Median	3Q	Max
-663,8	-67,3	-1,9	70,7	971,7

Coefficients:

	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 ***

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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 DF, p-value: < 2,2e-16

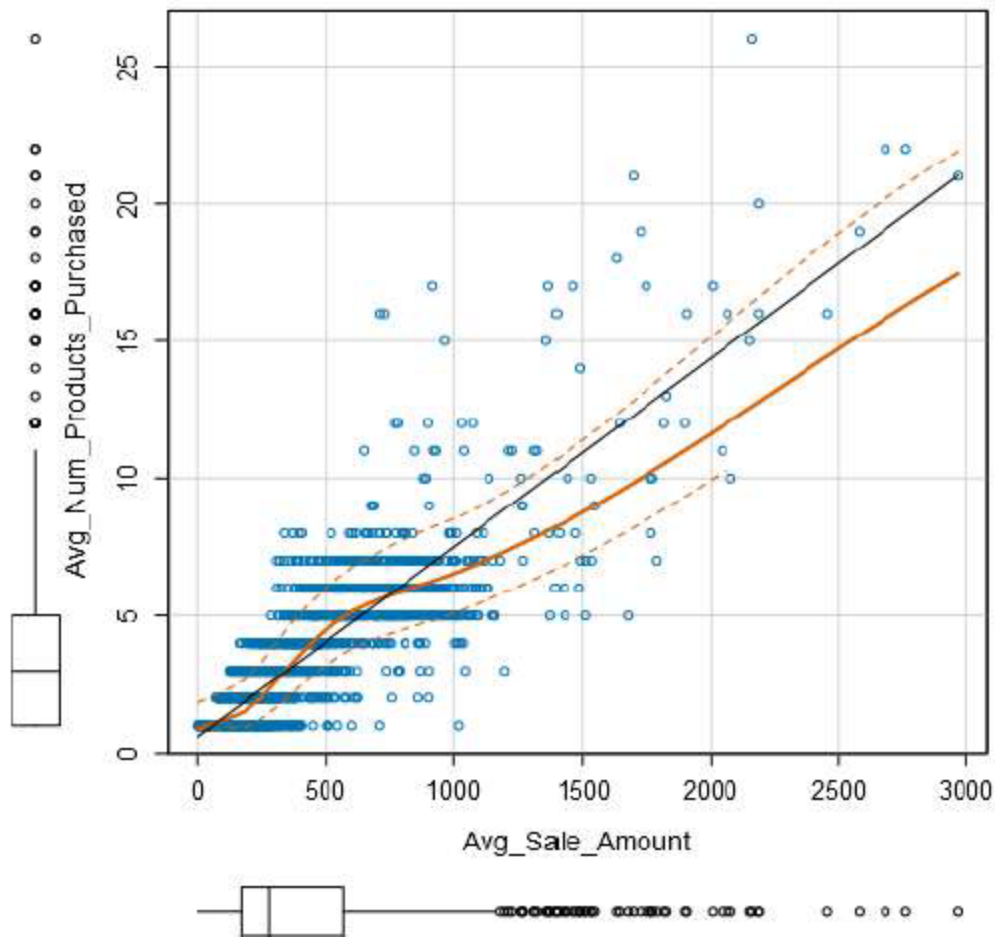
### *Type II ANOVA Analysis*

Response: Avg.Sale.Amount

	Sum Sq	DF	F value	Pr(>F)
Customer.Segment	28715078,96	3	506,4	< 2,2e-16 ***
Avg.Num.Products.Purchased	36939582,5	1	1954,31	< 2,2e-16 ***
Residuals	44796869,07	2370		

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

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.

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.

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.

**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.

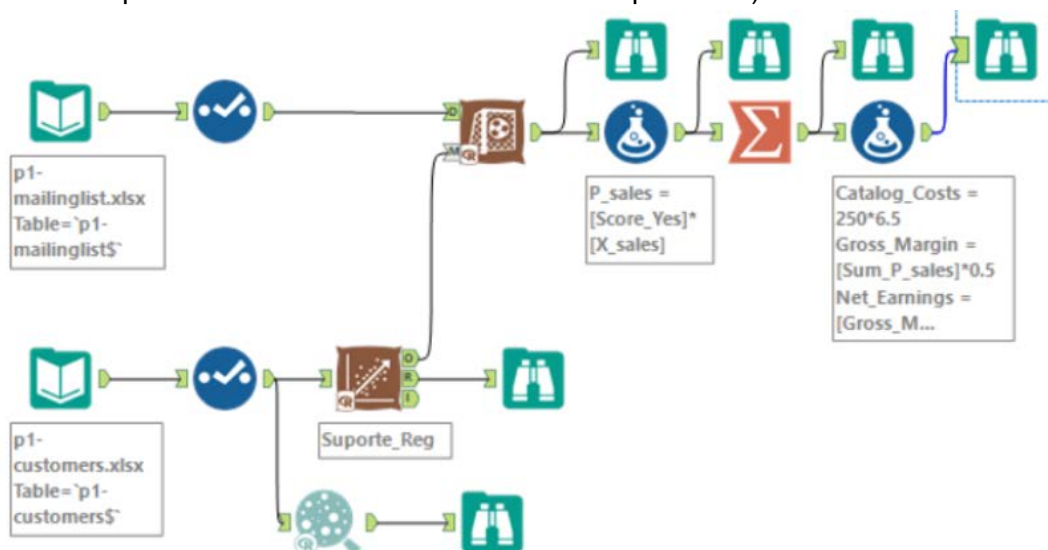
$Y = 303.46 + (281.84 \times \text{Customer\_SegmentLoyalty Club and Credit Card}) + (-149.36 \times \text{Customer\_SegmentLoyalty Club Card Only}) + (-245.42 \times \text{Customer\_SegmentStore Mailing List}) + (66.98 \times \text{Avg\_Num\_Products\_Purchased}) + \text{Credit Card} \times 0$

### 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.
2. 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 previsão 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.

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

Record #	Sum_P_sales	Catalog_Costs	Gross_Margin	Net_Earnings
1	47224.871373	1625	23612.435687	21987.435687

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](#). Os revisores vão usar esta rubrica para avaliar o seu trabalho.