

MAKE A COPY

Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Provide an explanation of the key decisions that need to be made. (500 word limit)

Key Decisions:

Answer these questions

1. What decisions need to be made?

We need to decide whether or not the company will have profit that exceeds \$10,000 from sending a catalog to the 250 customers from the mailinglist and decide to send catalogs or not.

2. What data is needed to inform those decisions?

Dataset about the previous customer already provided by the company, the data for the new customer, the cost and average gross margin and the cost of printing.

Step 2: Analysis, Modeling, and Validation

Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

1. How and why did you select the predictor variables in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer back to the "Multiple Linear Regression with Excel" lesson to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.

2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that your model produced.

3. What is the best linear regression equation based on the available data? Each coefficient should have no more than 2 digits after the decimal (ex: 1.28)

Important: The regression equation should be in the form:

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

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

Note that we **must** include the 0 coefficient for the type Cash.

Note: For students using software other than Alteryx, if you decide to use Customer Segment as one of your predictor variables, please set the base case to Credit Card Only.

I have used average sale amount as target variable and the predictor variable as following:

- Customer_Segment
- Avg_Num_Products_Purchased

Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366

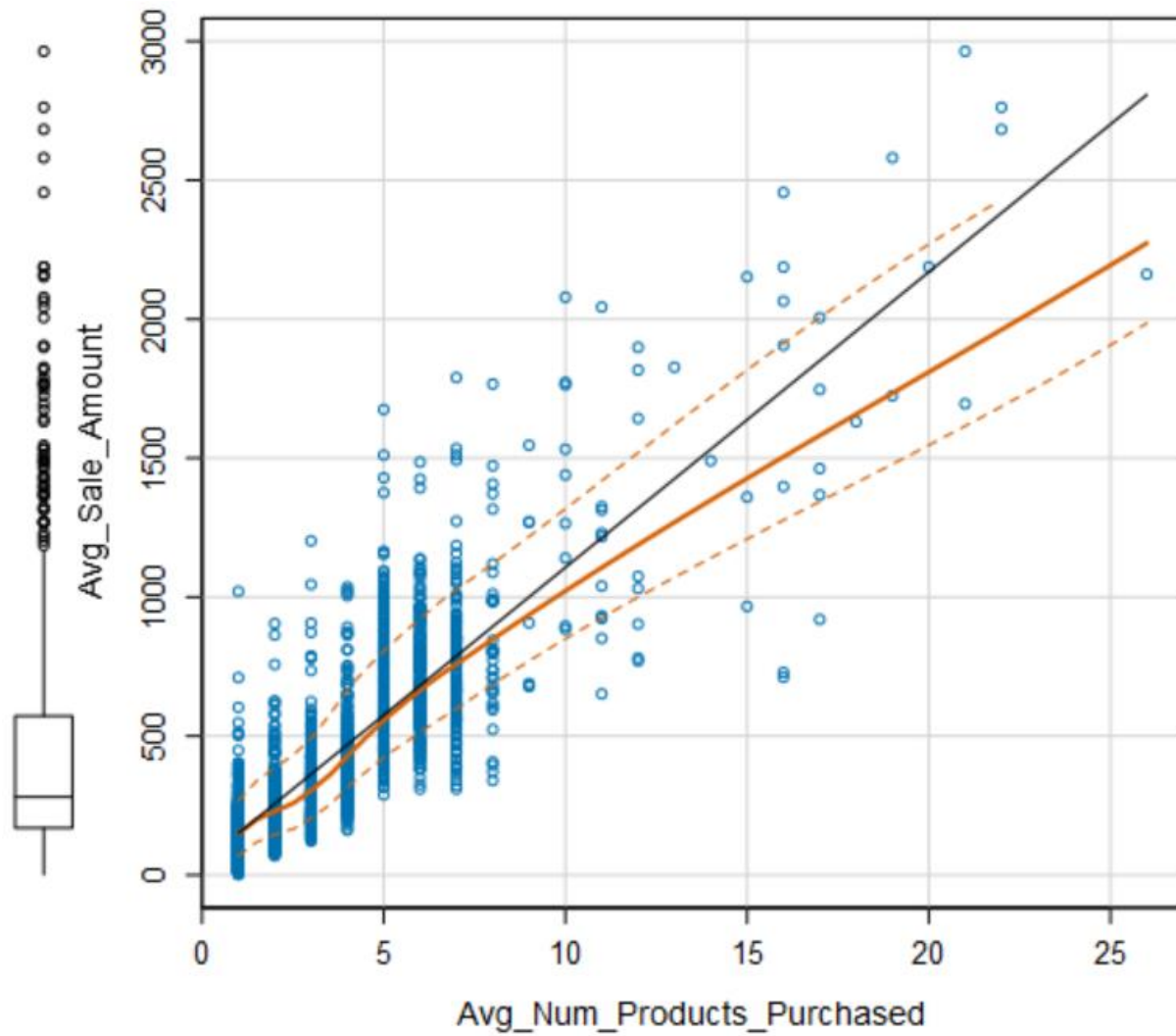
	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	303.46	10.576	28.69	< 2.2e-16	***
Customer_SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16	***
Customer_SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16	***
Customer_SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16	***
Avg_Num_Products_Purchased	66.98	1.515	44.21	< 2.2e-16	***

From above statistical outcome all variables has p-value less than 0.05 and 3 stars, which indicates all variables are significant. In addition, R-squared nearly 84%

This is the best linear regression equation

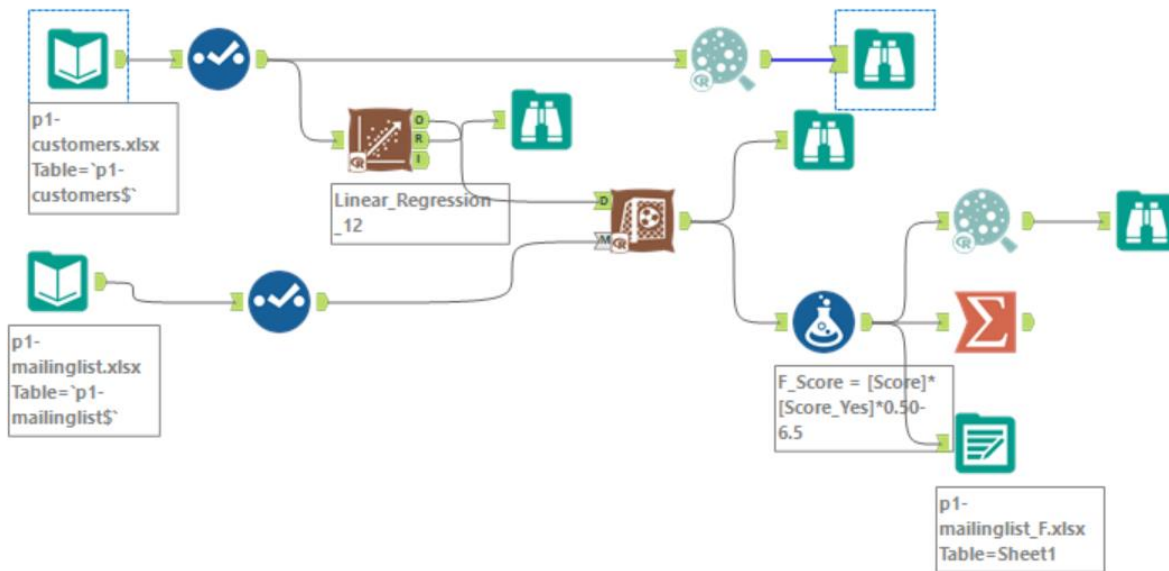
$\text{Avg_Sale_Amount} = 303.46 - 149.36 * (\text{Customer_SegmentLoyalty Club Only}) + 281.84 * (\text{Customer_SegmentLoyalty Club and Credit Card}) - 245.42 * (\text{Customer_SegmentStore Mailing List}) + 66.98 * (\text{Avg_Num_Products_Purchased})$

We see from below scatterplot that there is a linear relation between average sale amount and average number products purchased.



Step 3: Presentation/Visualization

Use your model results to provide a recommendation. (500 word limit)



At the minimum, answer these questions:

1. What is your recommendation? Should the company send the catalog to these 250 customers?

Yes

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)

I applied the model on the mailinglist data to predict the average sales amount, then I multiplied the average sales amount by the probability of customers which will buy to get the revenue, then I multiply the revenue by 0.5 to get the average gross margin. Finally, subtracting 6.5 (costs of printing and distributing) to get the profit.

3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?

21,987.44 which is more than 10,000

Before you Submit

Please check your answers against the requirements of the project dictated by the [rubric](#) here. Reviewers will use this rubric to grade your project.