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 needs to be made?

Key business decisions need to be made. The task is to determine the expected profit from 250 new customers and that the expected profit contribution should exceed $10,000.

2. What data is needed to inform those decisions?

The data needed to inform these decisions are a dataset from last years customers to build the model upon and get insight on past trends to make predictions. The second dataset is from new clients that will be used to predict sales estimate how much revenue the company can expect if they send out the catalog.

Step 2: Analysis, Modeling, and Validation

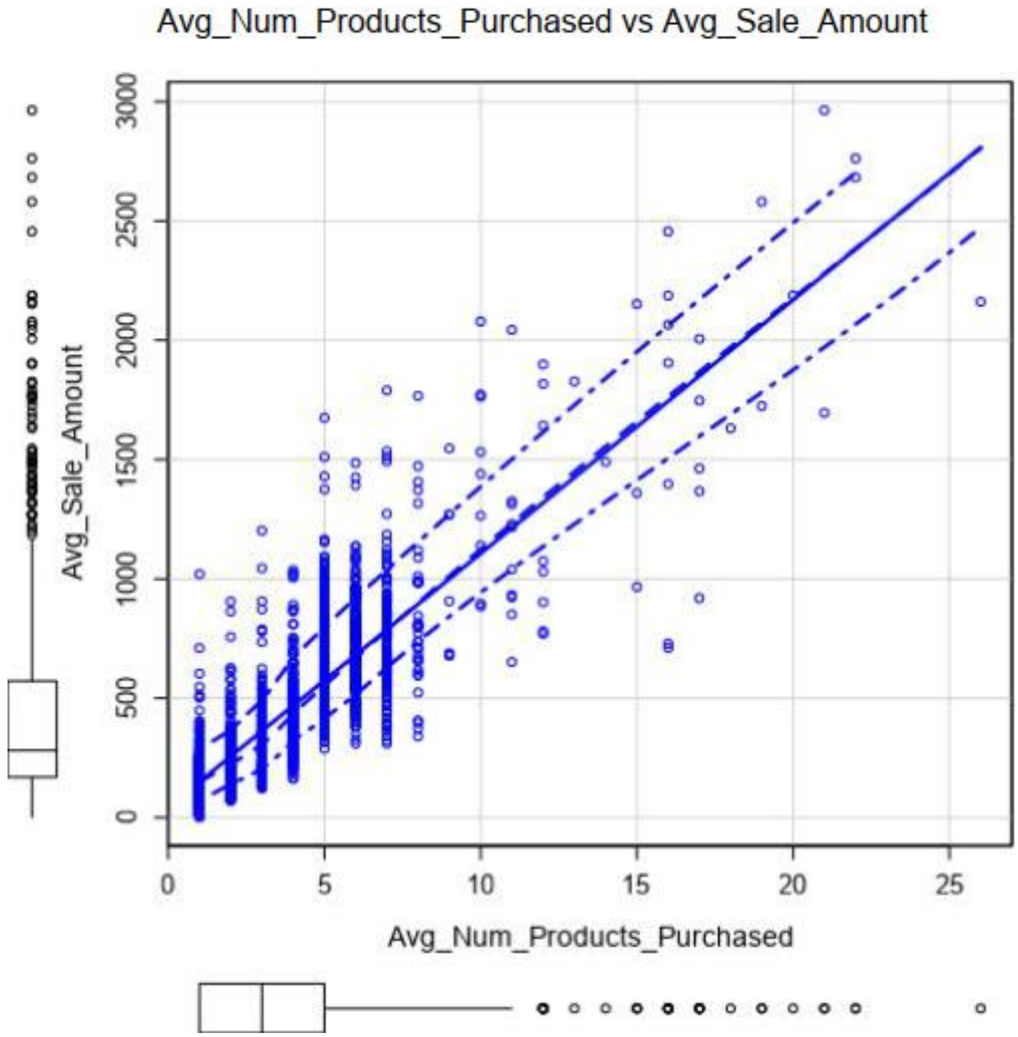
Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

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.

Once the data has been uploaded I created scatterplots for all numeric variables to see where the relationship was strongest. It was Avg_Num_Products_Purchased and Avg_Sale_Amount. I also used Dummy variables instead of Responded_to_Last_Catalog, but it showed no relationship with Avg_Sale_Amount.



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.

The p-values < 2.2e-16, which means the hypothesis is true. (p-value(probability value) is a quantitative measure to report the result of statistical hypothesis testing.) The p-value was the same for Customer_Segment and Avg_Num_Products_Purchased.
R-squared 0.84 (R-squared explains to what extent the variance of one variable explains the variance of the second variable) which in this case means that approximately 84% of the observed variation can be explained by the model's inputs.

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:

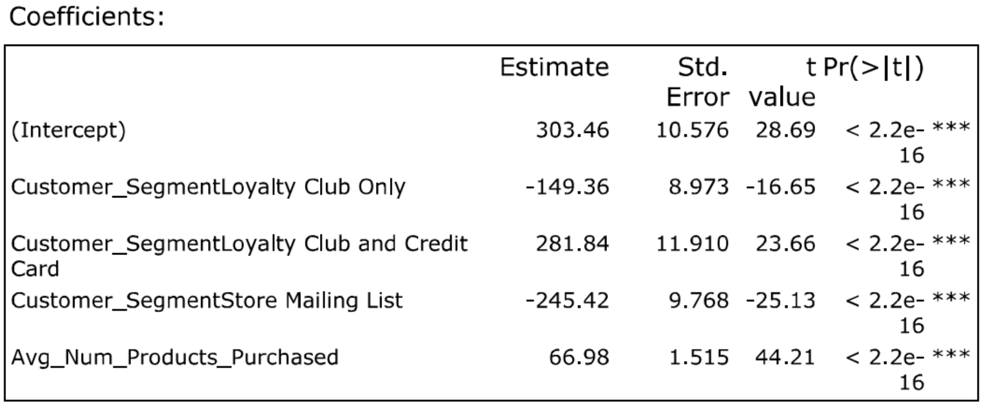
*Y = Intercept + b1 * Variable_1 + b2 * Variable_2 + b3 * Variable_3……*

For example: Y = 482.24 + 28.83 * Loan_Status – 159 * Income + 49 (If Type: Credit Card) – 90 (If Type: Mortgage) + 0 (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.

Y = 303.46 + (-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



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	***

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?

The sum expected revenue when calculated is $21,987.4356 USD. According to the minimum revenue expected from the company, this is a very good indicator that the company should send the catalogs to the new customers.

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

- First I built a Regression model with the dataset containing information about previous year sales and customers. I used the variables that showed a linear correlation. The output was then processed in the Score tool with the data from the new customers in order to creates an estimate of a target variable to a set of supplied predictor variables. The output is then used in the formula "([Score]*[Score_Yes]*0.5)-6.5" so that it calculates profit for each of the new customers, taking in consideration that The costs of printing and distributing is $6.50 per catalog and the average gross margin (price - cost) on all products sold through the catalog is 50%. In the end the sum of those results was calculated to gain the value of the expected profit.

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

The expected profit from the new catalog is $21,987.4356 USD

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.