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?

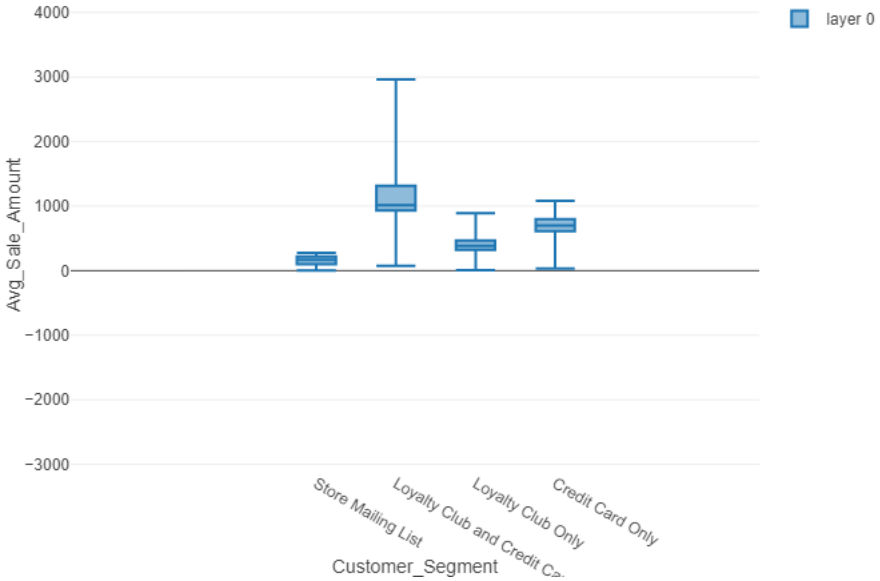
The company ultimately trying to make a decision on whether we should send the catalog to the new 250 customer in the mail list. The criteria is to send if expected profit contribution exceeds $10,000. We are going to build a regression model to predict expected profit.

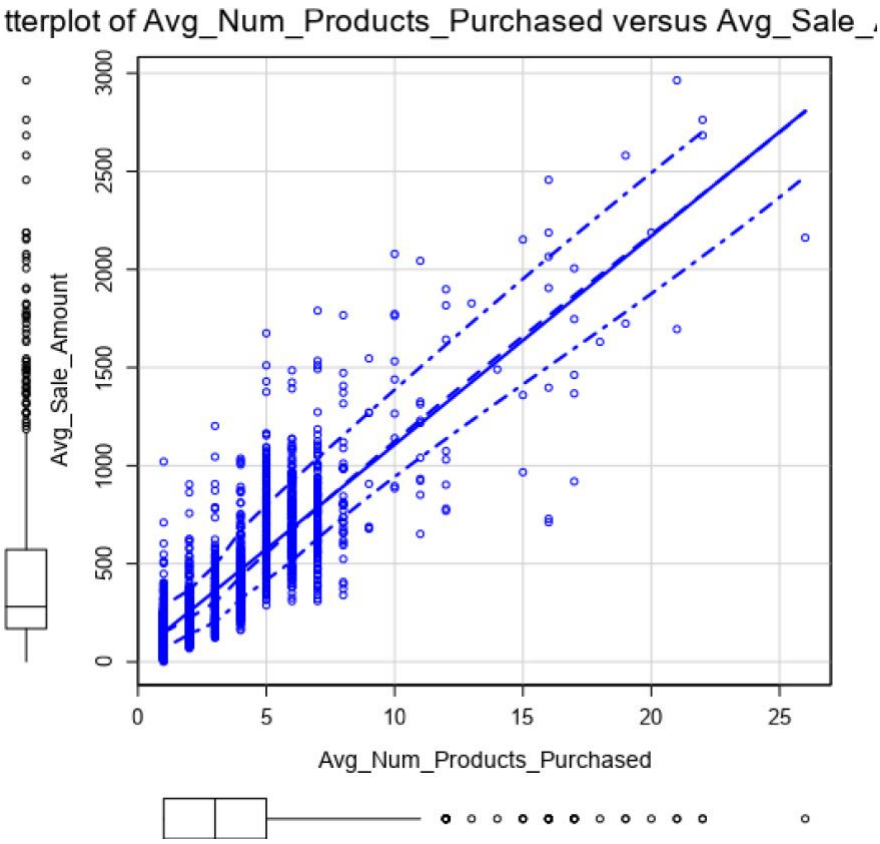
2. What data is needed to inform those decisions?
Previous customer characteristics and the number of their average sale amount.

Step 2: Analysis, Modeling, and Validation

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.

We choose to use customer segment, and number years as customer. We can see from the scatterplot and boxplot that those two variables are correlated with average sales. Although we considering using average number product purchased as well, the scatterplot shows that the correlation is weak.





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.

Our model have a large R-squared of 0.8369 and Adjusted R-Squared of 0.8366. Also, every single predictor variables are significant (p <= 0.05)

	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 degrees of freedom (DF), p-value < 2.2e-16

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)

Y = 303.46 + 66.98 * avg_num_products_purchased – 149.36* (If: Loyalty Club Only) – 245.42* (If Store Mailing List) + 281.84* (If Loyalty Club and Credit Card) + 0 (If Credit Card Only)

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?

We recommend the customer to send the catalog to these 250 customers.

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

We use regression model to calculate the average sale (score) for each customer in the mailing list. Then, we multiply it by “Score_Yes” to calculate the potential revenue. Then, we calculate the expected profit for each person by multiplying potential revenue and 0.5(gross margin) and then minus 6.5 (costs of printing and distributing). Then, we sum expected profit for each person to get the expected profit from the new catalog.

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

The potential profit is 21987.435687.

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.