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 decision that needs to be made is whether to send catalog to the new 250 customers based on calculated profit.

2. What data is needed to inform those decisions?

   a. Customer segment
   b. Average number of products purchased
   c. Score yes
   d. Margin
   e. Cost of catalog

Step 2: Analysis, Modeling, and Validation

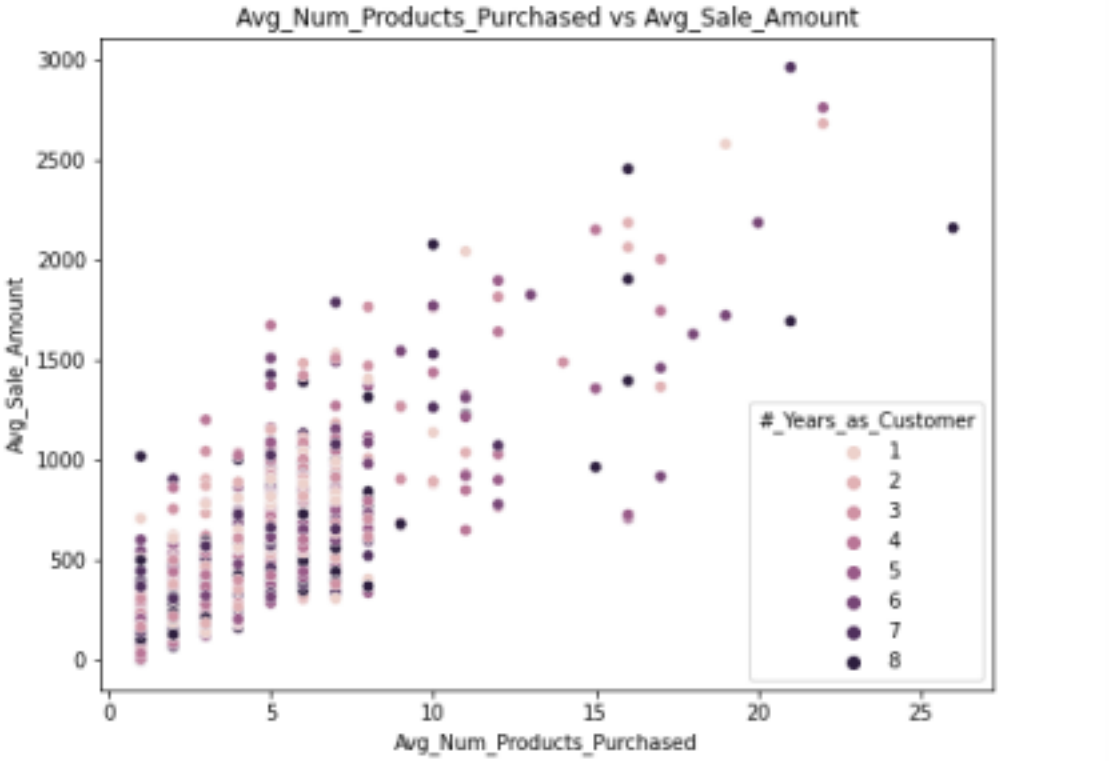
Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

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.

I did plot a scatter plot of Avg_Num_Products_Purchased vs 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.

P- value < 2.2e-16
Adjusted R-squared = 0.83
This suggest our model is good because it has p- value less that 0.05 and adjusted R-squared more than 0.7

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 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)

a= Customer_SegmentLoyalty Club Only
b= Customer_SegmentLoyalty Club and Credit Card
c= Customer_SegmentStore Mailing List
d= Avg_Num_Products_Purchased

Y = 303.46 - 149.36 * (a) + 281.84 * (b) - 245.42 (c) + 66.98 (d)

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?

Send the catalog to 250 customers.

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

I calculated the predicted sale first. After that I calculated the actual sale by multiplying the predicted sale * yes score. Then calculated as done below on number 3

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

Since Expected profit = (Sum of expected revenue * Gross margin) - (Cost of catalog * 250) =

(47225.87 * 0.5) - (6.50 * 250)

= $ 21987.44