<u>Project 1: Predicting Catalog Demand</u>

Step 1: Business and Data Understanding

Key Decisions:

Answer these questions

1. What decisions needs to be made?

A company that manufactures and sells high-end household goods. the company sent out its first print catalog last year and is planning to send out the catalog for this year in the coming months. The business has 250 potential clients who they want to send the catalog to from their mailing list. So the company want to predict the profit expected from these 250 new customers but they does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000.

2. What data is needed to inform those decisions?

We need last year sales data and processing it to get insight from to predict the outcome for the decisions we try to make, also we need information about the cost of shipping and preparation to calculate profit.

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.

There is variables does not have relationship with the outcome such as customer name or id because they don't change average sales, So the linear relationship is between Avg_sale_Amount and Avg_Num_Products_purchased. I include categorical variables in the regression model and see if the coefficients are significant increase the adjusted r. because if there is a linear relationship, then coefficients are significant and the r must be relatively high.

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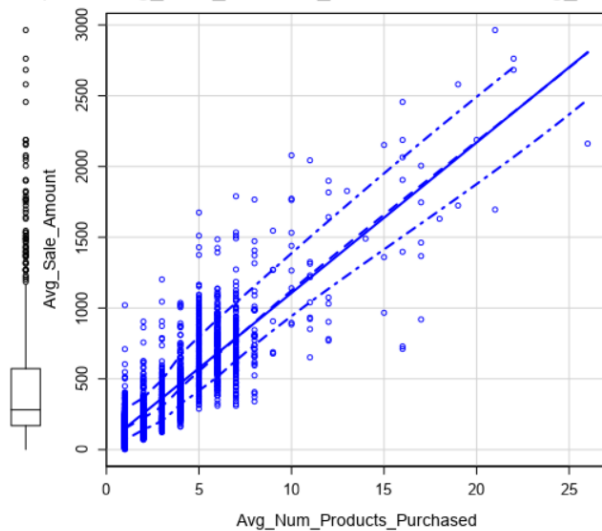
2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that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

I believe the model is a good model because the R-squared has a good result which is above .80

Record	Report
1	**Report for Linear Model Linear_Regression_9**
2	*Basic Summary*
3	Call: lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)
4	Residuals:
5	

	Min	1Q	Median	3Q	Max
	-663.8	-67.3	-1.9	70.7	971.7

Record	
6	Coefficients:
7	

	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

Record	
8	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
9	*Type II ANOVA Analysis*
10	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

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)

Avg_Sale_Amount = 303.46 + 66.98 * Avg_Num_Products_Purchased -149.36 (If Customer_Segment: Loyalty Club Only) + 281.84 (If Customer_Segment is Loyalty Club and Credit Card) – 245.42 (If Customer_Segment is Store Mailing List) + 0 (If Customer_Segment is Credit Card Only)

Step 3: Presentation/Visualization

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

   The expected profit exceeds the $10.000 so I will recommend sending catalog to new customer

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

   Calculated avg_sale_amount using linear model, and the probability if the customer will respond to catalog or no, and also calculated the profit [predicted sum*0.5-(6.50*250)] 250 number of customer and 6.50 the price of the catalog

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

 $21987.43

Workflow

p1-
customers.xlsx
Query=`p1-
customers$`

Linear_Regression
_9

p1-
mailinglist.xlsx
Query=`p1-
mailinglist$`

Predicted =
[Score]*
[Score_Yes]

Predicted_value =
[Sum_Predicted]
*0.5-(6.50*25)