Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

.

1. What decisions needs to be made?

The decision needs to be made is, “Should we send the Catalog to the new customers?”

2. What data is needed to inform those decisions?

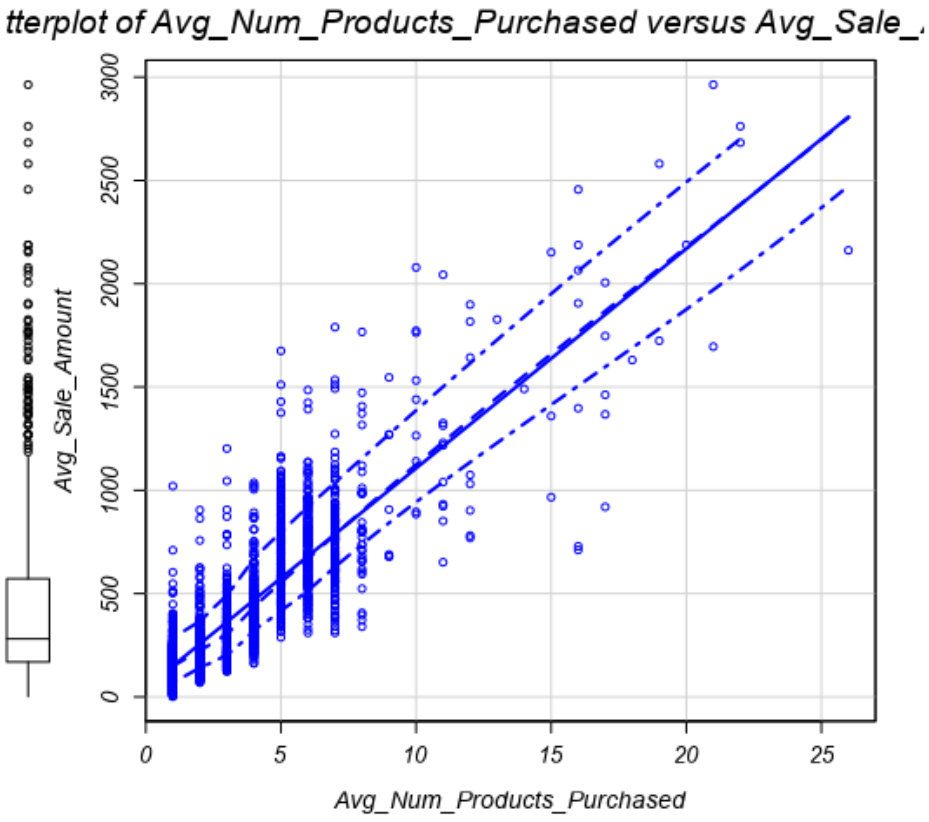
Data needed to inform sales predictions and expected profits are: Customer_Segment , Avg_Num_Products_Purchased , Store_Number , #_Years_as_Customer , Avg_Sale_Amount , Gross Margin, Cost of catalog printing & distribution to 250 customers

Step 2: Analysis, Modeling, and Validation

I have selected the predictor variable in my model by using the scatterplot to find out whether or not there is a good relationship between the predictor variables and the target variable.

For this project, we need to calculate expected profit. We can do this by predicting Average Sales, and then using the cost and profit margins to come up with the expected profit. So, our **Target Variable** will be *Average Sales*. In a scatterplot, the Target Variable is represented in the y axis.

All other variables in the ``p1-Customer`` file are potential predictor variables. The predictor variables go on the x axis (horizontal axis of the scatterplot).



The linear model is a good model because our statistical result shows a multiple R-squared value
of 0.8369 and adjusted R-squared value of 0.8366---both of which are high value and greater than 0.70. A high r-squared (0 to 1) indicates a greater explanatory power of the model and represents the amount of variation in the target variable associated with the variation in the predictor variables. Thus, a 0.8366 adjusted R-squared shows that 83.66% of the variance can be
explained by the model.
It's also a good model since the p-value for Customer Segment and Average Number of Products
is less than 0.05 showing their statistical significance. P-value indicates the probability that the coefficient is zero.

Report for Linear Model Catalog_Regression

Basic Summary

Call:
lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)

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

Type II ANOVA Analysis

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

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

Step 3: Presentation/Visualization

The company should send the catalog to these 250 customers since the catalog is predicted to generate a profit of $21,987.44 after factoring the costs and margins---and exceeds the $10,000 expected profit contribution. See calculation in question 3. The expected revenue from each customer is calculated by multiplying the X field (predictedsales amount) with their respective Score_Yes value (probability of purchasing products).
Summing the expected revenues for all 250 customers generates a total of $47,224.87. Next, the
sum of expected revenue for all 250 customers is multiplied by 0.50 to account for the 50% gross
margin on all products sold through the catalog. Finally, the expected profit is calculated by subtracting (6.50*250) from the (summed expected revenue*gross margin).The expected profit of $21,987.44 exceeds the company's $10,000 contribution and thus, the company should send the catalog to the 250 customers.

Expected Profit = (Sum of expected revenue x Gross Margin) – (Cost of Catalog x 250)
= (47,224.87 x 0.5) – (6.50 x 250)
= 23,612.44 – 1,625
= $21,987.44