Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

1. Should the company send the mailing? The company has determined that they only want to send the mailing, if there is an expected $10,000 in profit after the cost of the mailing and accounting for a 50% gross margin on items ordered.

2. Which customers should receive the mailing? The current proposed mailing is 250 customers. To maximize profit, and reduce expenses should the mailing be sent to all 250 customers on the mailing list?

Data Needed:

We will examine information from our customer database including:
1. The customer segment
2. The Number of Products Purchased
3. The Number of Years as a Customer
4. The Average Sale for the Customer
5. If the Customer Responded to the Last Catalog
6. The Store Number

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.

I selected the predictor variables of my model, after loading the customers dataset into Alteryx and running a linear regression on the customer list. Based on the regression, the predictor variables in my model are the Customer Segment, and the Average Number of 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	***

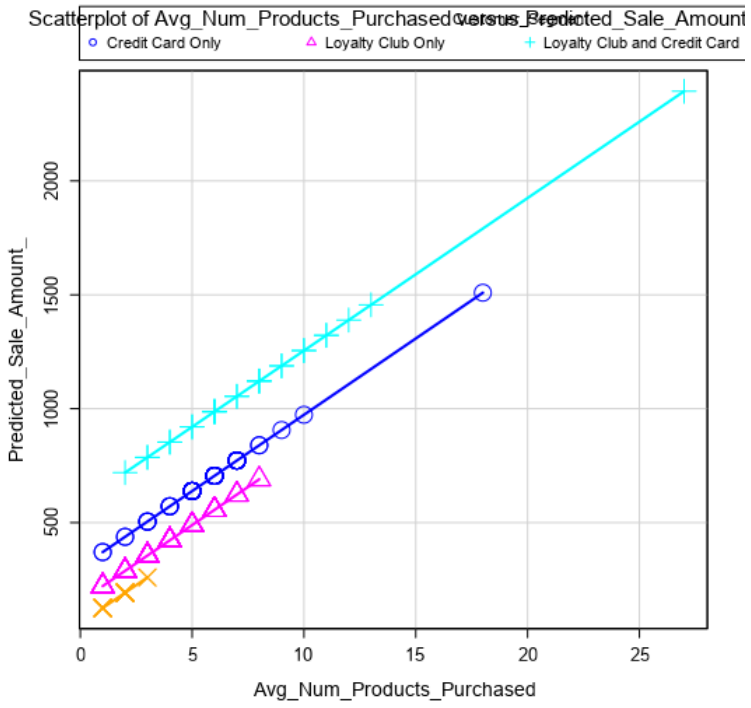
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

As you can see from the chart above for each variable selected, we can verify that they are good indicator variables based on their p-values being less than .001. The multiple R Squared value is 0.8369.

The best Linear Regression Equation based on this data is:

Y = 303.46 + 66.98 * Avg Products Purchased -149.36 * Customer Segment (If Type: Loyalty Club Only) + 281.84 * Customer Segment (If Type: Loyalty Club and Credit Card) -245.42 * Customer Segment (If Type: Store Mailing List)

Below you can visually see how the predicted sale amount is impacted by both customer segment and the average number of products purchased.



Step 3: Presentation/Visualization

Yes, based on our model the company could predict a profit of $21,987.96 by sending the mailing to all 250 customers.

This was calculated by applying our linear regression model to each customer. Then applying a Sum Product function in excel to the columns that predicated percent chance the customer would respond to the mailing (Score_Yes_Response_To_Mailing) by the predicted amount the customer would purchase (Predicted Sale_Amount). This new variable was saved as the "Expected Sales for Mailing". From there we could calculate the expected revenue, costs, and profit for mailing all 250 customers.

Inputs	Sum Product	$ L
Customers on Mailing List	250	
Minimum Profit for Mailing	$10,000.00	
Cost of Printing and Distrubuting	$ 6.50	
Gross Margin on ordered items	0.5	
Avereage % Expected to Order	0.34066812	
Expected Sales for Mailing	$47,225.91	
Expected Revenue for Mailing	$23,612.96	
Expected Cost for Mailing	$ 1,625.00	
Expected Profit for Mailing	$21,987.96	

With our regression formula we can further analyze the mailing list to determine if by segmenting the list there is an opportunity to generate greater profits, with better certainty.

Inputs	Sum Product	Send to 34% + Likely to Buy	Send to 50% + Likely to Buy	Segment Loyalty Club & Credit Card	Send to 50% & Segment
Customers on Mailing List	250	81	39	26	61
Minimum Profit for Mailing	$10,000.00	$10,000	$ 10,000.00	$ 10,000.00	$ 10,000.00
Cost of Printing and Distrubuting	$ 6.50	$ 6.50	$ 6.50	$ 6.50	$ 6.50
Gross Margin on ordered items	0.5	0.5	0.5	0.5	0.5
Avereage % Expected to Order	0.34066812	0.558793403	0.5	0.337051645	0.67865324
Expected Sales for Mailing	$47,225.91	$ 25,392.71	$ 10,834.02	$ 9,779.78	$ 17,888.53
Expected Revenue for Mailing	$23,612.96	$ 12,696.36	$ 5,417.01	$ 4,889.89	$ 8,944.27
Expected Cost for Mailing	$ 1,625.00	$ 526.50	$ 253.50	$ 169.00	$ 396.50
Expected Profit for Mailing	$21,987.96	$ 12,169.86	$ 5,163.51	$ 4,720.89	$ 8,547.77

Even though the average percentage of customers that responded to the mailing increased when we only sent the mailing to customers most likely to buy items from the catalog the overall profits in every scenario we ran were lower than if we were to send the mailing to everyone.

This is due to the relatively low cost of the mailing when compared to the average predicted sale amount per customer.

For example, by sending the mailing only to customers that have a likely to buy score 50% or greater, the average percentage of customers from the mailing expected to order increases to 55%. However, the number of customers the mailing is sent to decreases from 250 to 81, which has a larger impact on expected sales. Therefore, expected sales decrease from $47k to $25k, and overall expected profits decrease from $21k to $12k with the smaller mailing.

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.