Predicting Catalog Demand

Step 1: Business and Data Understanding

What decisions need to be made?

- This business needs insight on expected profit from 250 new customers. The insight helps shape the decisions on whether to send out catalogs or not. A predicted profit of over 10,000USD is expected before a decision to send out catalogs is made.

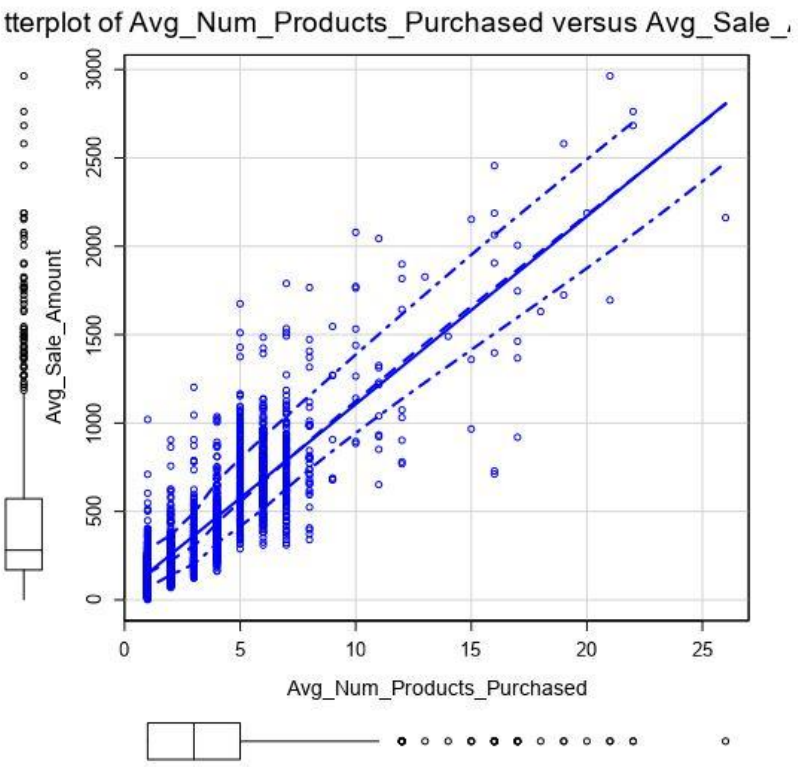
What data is needed to inform those decisions?

- To make this decision, we would be using:

    1. The historical data of past customers to build a model to predict the expected spending habits of the new customers.
    2. Data of the new customers on the mailing list is also needed to apply the prediction model to, so as to calculate the average sale per customer
    3. The average gross profit margin
    4. The cost of printing the catalogs

Step 2: Analysis, Modeling, and Validation

The average number of products (Avg_Num_Products) has a linear relationship with the average sale amount (Avg_Sale_Amount) so it's therefore a good predictor variable

Record	Graph
1	

Customer segment (Customer_Segment) is a categorical variable, therefore a scatterplot can not be created. However, the multiple-R-squared is high and the coefficients are significant therefore there is a linear relationship

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

A prediction model was built using the fields *Customer_Segment*, *and Avg_Num_Products_Purchased* as the predictor variables with *Avg_Sale_Amount* as the target variable.

p-values and R-squared values

The closer an R-squared value to 1 (0.7 and above) the better the model. This model has an adjusted R-square of **0.8366**, therefore the two predictors make a strong model.

With a p-value of less than $\mathbf{2.2e^{-16}}$ which is smaller than the threshold 0.05. The relationship between the predictor variables and the target variable is considered statistically significant

Linear regression equation

The tool used for the analysis has automatically transformed the categorical variables into dummy variables therefore giving us the equation below:

*Sale amount = 303.46 + (-149.36 x **loyalty club only**) + (281.84 x **Loyalty Club and Credit Card**) + (-245.42 x **Store Mailing List**) + 66.98 (**Avg_Num_Products_Purchased**)*

*Sale amount = 303.46 - (149.36 x **loyalty club only**) + (281.84 x **Loyalty Club and Credit Card**) - (245.42 x **Store Mailing List**) + 66.98 (**Avg_Num_Products_Purchased**)*

Step 3: Make a recommendation

My recommendation would be that my manager goes ahead to send those catalogs because a profit of $21,978.44 is predicted to come into the company after sending them out.

After building the model, I applied the score tool to get the sum of projected profit my model predicted. This came to a total of $47,224.87. This sum is inclusive of the company's gross margin.

To get the net profit, I had to deduct the margin and cost of printing the catalogs. I multiplied the projected profit by 0.5 which represents the 50% margin, and multiplied $6.50 (cost of printing one catalog) by 250 (number of new customers). The formula can be found below

*[Sum_Projectedprofit] - ([Sum_Projectedprofit] *0.5) - (6.5*250) = $21,987.44*

This brings the expected net profit above $10,000.

However, it is worthy of note that from the past data, sales have not come from sending out these catalogs. Out of over 2000 customers, 92.8% didn’t respond to the last catalog sent out and that didn’t affect the sales. I would ask my manager that they provide more data to be able to look into what drives sales in the company.