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

Step 1: Business and Data Understanding

1. What decisions needs to be made?

A company that manufactures and sells high-end household goods and is preparing to send a catalog this year in the coming months and the company has 250 new customers from the mailing list to which they wish to send the catalog.
The company's management wants to determine whether the print catalog should be sent to 250 new customers in the coming months. If the expected profit contribution from these new customers exceeds $10,000, the company will send the catalog. Less than the expected profit was the company would not send the catalog.

2. What data is needed to inform those decisions?
 - You should look at last year's sales data for a company and know how likely it is for the new customer to buy or not.
- Identify information about customers and their shopping behavior.
- Processing the cost structure and determining the profit margin.
- Predict the probability the new customer will buy or not to get the sales data for the current year and then decide to send the catalog or not.

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.

The Name and Customer_Id do not depend on sales because they do not change the average sales so are data that is not used by analyzing the data as the address is considered very detailed and can be dispensed with and use the zip code.

Used Analyzed the scatterplot in alteryx
there is a linear relationship between Avg_Num_Products_Purchased and Avg_Sale_Amount.

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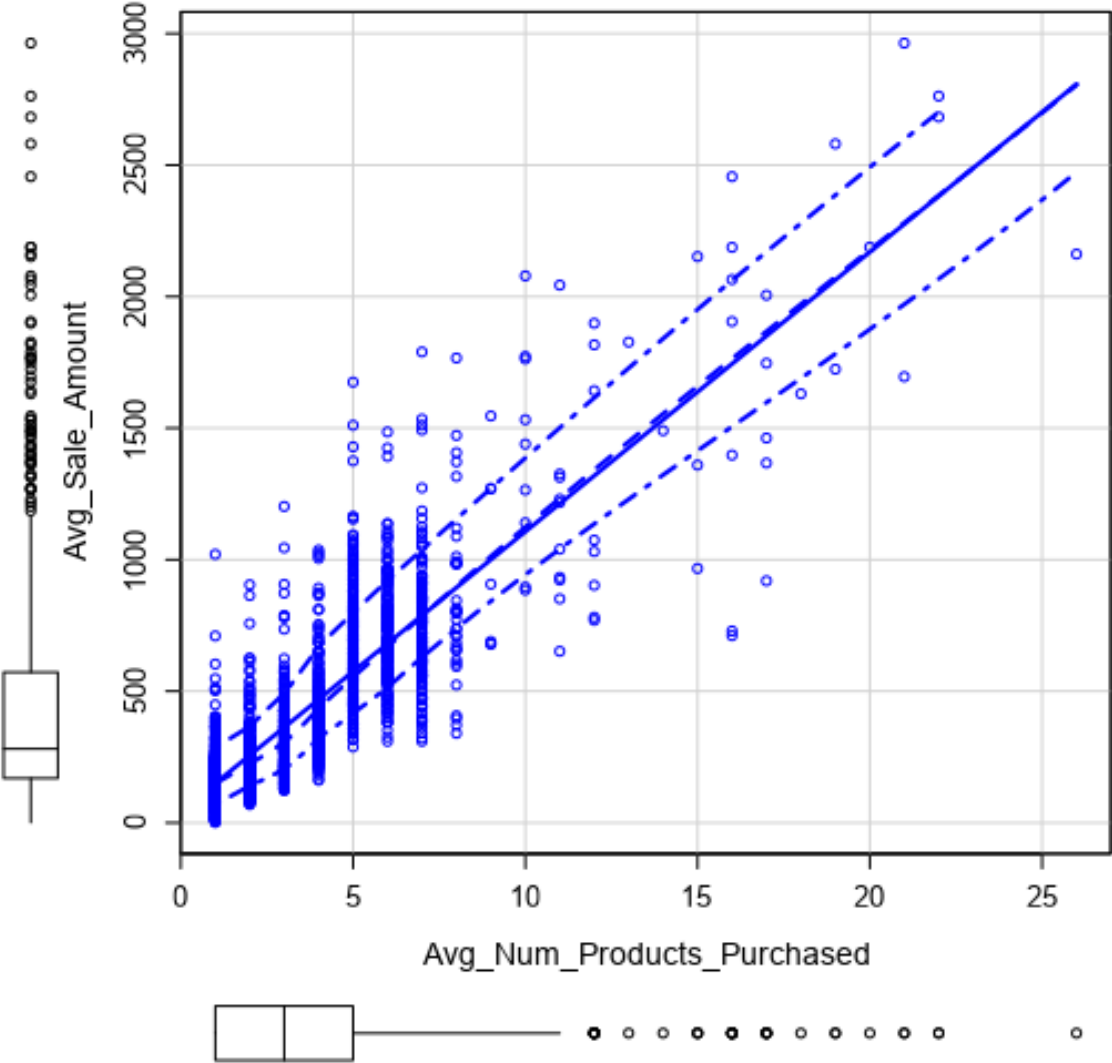
2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that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

Record	Report
1	**Report for Linear Model Linear_Regression_29**
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

This model is strong since the R-value is very high (0.8366).

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?

If the profit exceeds $10,000, the company will be dispatched and after using the linear regression model exceeded expectations.
Yes, the company must send the catalog to new customers.

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

1/Calculated Avg_Sales_Amount using linear regression model.
2/ Score_Yes : Probability that a customer will respond to catalog and make a purchase
3/Created a new column (X = Avg_Sales * Score_Yes) Given profit margin is 50%, and cost for each catalog is $6.50.
4/Calculated the profit = X 0.5-(6.50*250)

3.   What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?
 Profit equals $21987.43

Alteryx workflow:

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

Linear_Regression
_29

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

X =
[Average_Sales]*
[Score_Yes]

Sum_X = [Sum_X]
*.5-(6.50*250)