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

Step 1: Business and Data Understanding

Provide an explanation of the key decisions that need to be made. (500 word limit)

Determine if sending the catalog to 250 new customers will generate a profit that exceeds $10,000. Else, do not send the catalog to them.

Key Decisions:

Answer these questions

1. What decisions needs to be made?
   If the prediction for profit is greater than $10,000 then send the catalog to the 250 new customers. Else, do not send the catalog.

2. What data is needed to inform those decisions?
- *Historical data that can be used to generate a linear model to predict sales.*
- *The data for the new 250 customers with the same predictor variables used in the model, so can be predicted the sales and profit.*

Step 2: Analysis, Modeling, and Validation

Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

1. How and why did you select the predictor variables in your model?

Take for instance the AVG_Num_Products_Purchased has a linear relationship with the Avg_Sales, because as the value increase for AVG_Num_Products_Purchased, the Avg_Sales also increase with a linear pattern.

erplot of Avg_Num_Products_Purchased versus Predicted_A'
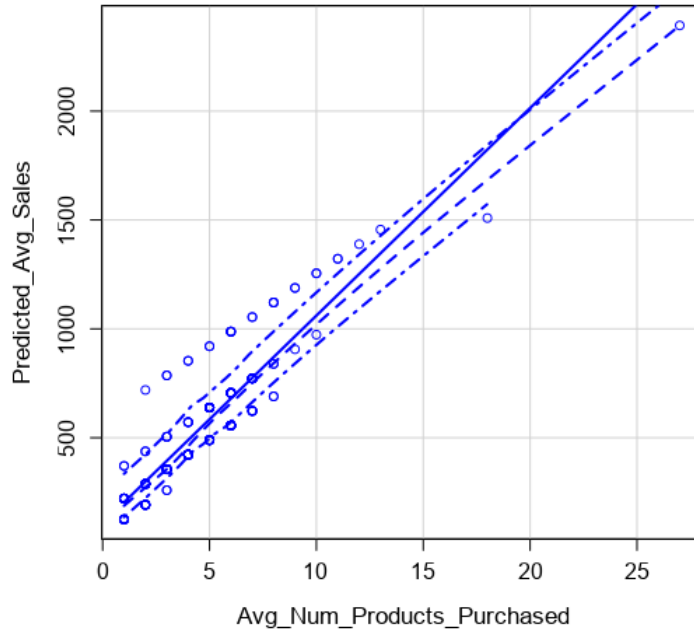
Also the Years_as_Customer was evaluated, but the scatter plot did not show a linear relationship with the target variable Avg_Sales_Amount.

Scatterplot of Years_as_Customer versus Avg_Sale_Amou
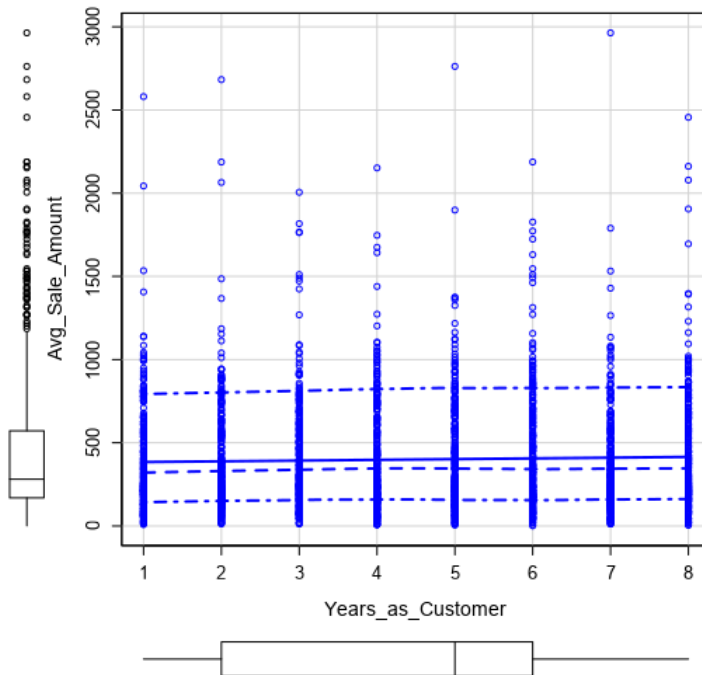
Other numeric fields like zip code, customer_id, store_number are not continue variables. Therefore, were not taken in consideration for plotting.

2. Explain why you believe your linear model is a good model.

The linear model has a good level of accuracy per the statistical elements generated by the Alteryx linear regression report.

The Customer_Segment and AVG_Num_Products_Purchased were selected as predictor variables, because per the linear model report, these variables have a probability of 2.2e-16, which is lower than 0.05, and that is a good indicator that are related with the target variable.

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	303.46	10.576	28.69	< 2.2e-16	***
Avg_Num_Products_Purchased	66.98	1.515	44.21	< 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	***

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

Also, because it is a multiple linear regression model, it was considered the value for Adjusted R-Squared: 0.8366, which is greater than 0.7, and good indicator of the model accuracy to predict the target variable.

3. What is the best linear regression equation based on the available data?

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

Step 3: Presentation/Visualization

Use your model results to provide a recommendation. (500 word limit)

At the minimum, answer these questions:

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

Yes, it is recommended to send the catalog to the new 250 customers, because the predicted profit is $21,987.43, considering that minimum required was $10,000, and the predicted profit is more than twice.

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

After running the model and apply the probability for purchase the results are:

- Revenue: $47,224.87
- Cost: $23,612.44 (because the gross margin is 50 %, therefore the Cost = $47,224 - (0.50 * $47,224))
- Total Cost for catalog print and distribution: $1,625 (250 new customers * 6.5 price unit)
- Profit = $47,224.87 - $23,612.44 - $1,625 = $21,987.43

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

$21,987.43