<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)

Key Decisions:

Answer these questions

1. What decisions needs to be made?
   A company needs to decide whether to send catalog to the new 250 customers in the mailing list. The passing criteria for this is that the Company needs the model to predict an expected profit contribution of above $10,000.

2. What data is needed to inform those decisions?
   Average number of products purchased
   Customer Segments.

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)
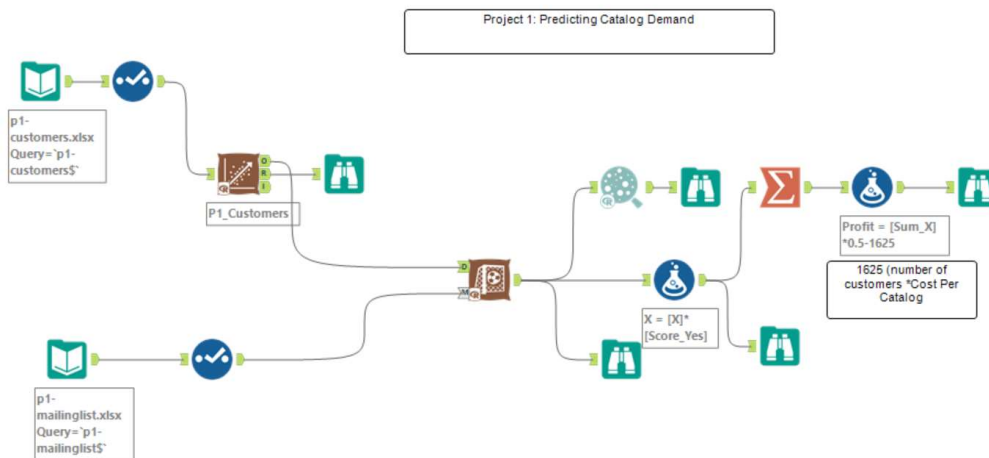
I trained the regression model with P1-customer data given. I used the linear regression model to predict (Average Sale Amount), I trained the model with (customer segment and average number of products purchased). I used the score model to compare both P1_customer and P1_mailing list. Then I predicted the expected revenue (Average sale amount * Score_Yes which is the

probability the customer will respond). After finding the sum I considered the gross margin and the cost of sending a catalog.

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? 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.
X signifies the Predicted the Average Sales Amount.

Scatterplot of Avg_Num_Products_Purchased versus X
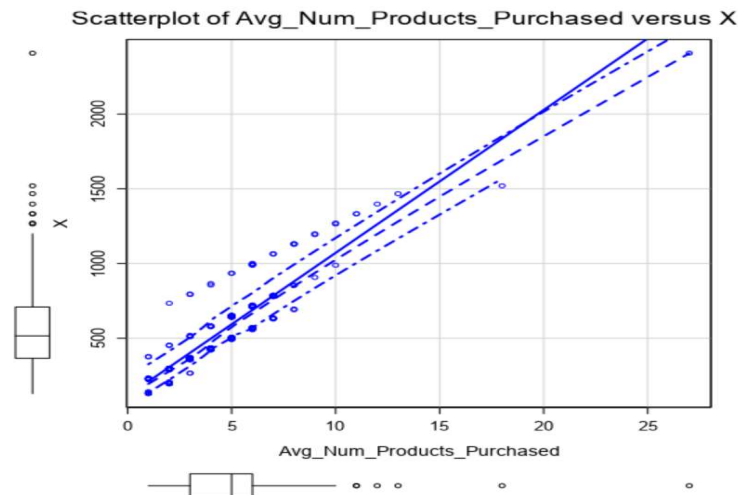
There is a linear relationship between Average Number of Products produced and the Average Predicted Sales Amount. The plot shows a higher Predicted Average Sales Amount at a higher Number of products Purchased.

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.

Report for Linear Model P1_Customers

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

The predictor variables Customer Segment and Average Number of products purchased P values have the highest significance of (***).
The prediction has an adjust R value of 0.8369 which shows it is a good model.

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)
X = Predicted Average Sales Amount
X = 303.46 – 149.36 *(Customer_SegmentLoyalty Club Only) +281.84*(Customer_SegmentLoyalty Club and Credit) -245.42*(Customer_segmentStore Mailing List) + 0 *(Customer_segmentCredit Card Only) + 66.98 * (Avg_Num_Products_purchased)

Important: The regression equation should be in the form:

Y = Intercept + b1 * Variable_1 + b2 * Variable_2 + b3 * Variable_3……

For example: Y = 482.24 + 28.83 * Loan_Status – 159 * Income + 49 (If Type: Credit Card) – 90 (If Type: Mortgage) + 0 (If Type: Cash)

Note that we **must** include the 0 coefficient for the type Cash.

Note: For students using software other than Alteryx, if you decide to use Customer Segment as one of your predictor variables, please set the base case to 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?
   The company can send the catalogs because the expected profit exceeds the pass criteria of $10,000.

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)
   I came up with the recommendation because the expected profit far exceeds the criteria of $10,000 and our model has a great P and R values. Therefore, the expected profits would not be too far from the predicted profits.

3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?
   The expected Profit from the new 250 customers is approximately $22,000.

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.