Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

What we need to determine here is if sending the catalogs to the 250 new customers is worth it profitwise.
Since the manager wants to send the catalogs only if the profit contribution exceeds $10,000 we have to calculate just that.

Key Decisions:

1. What decisions needs to be made?

If we will end up sending the catalogs.

2. What data is needed to inform those decisions?

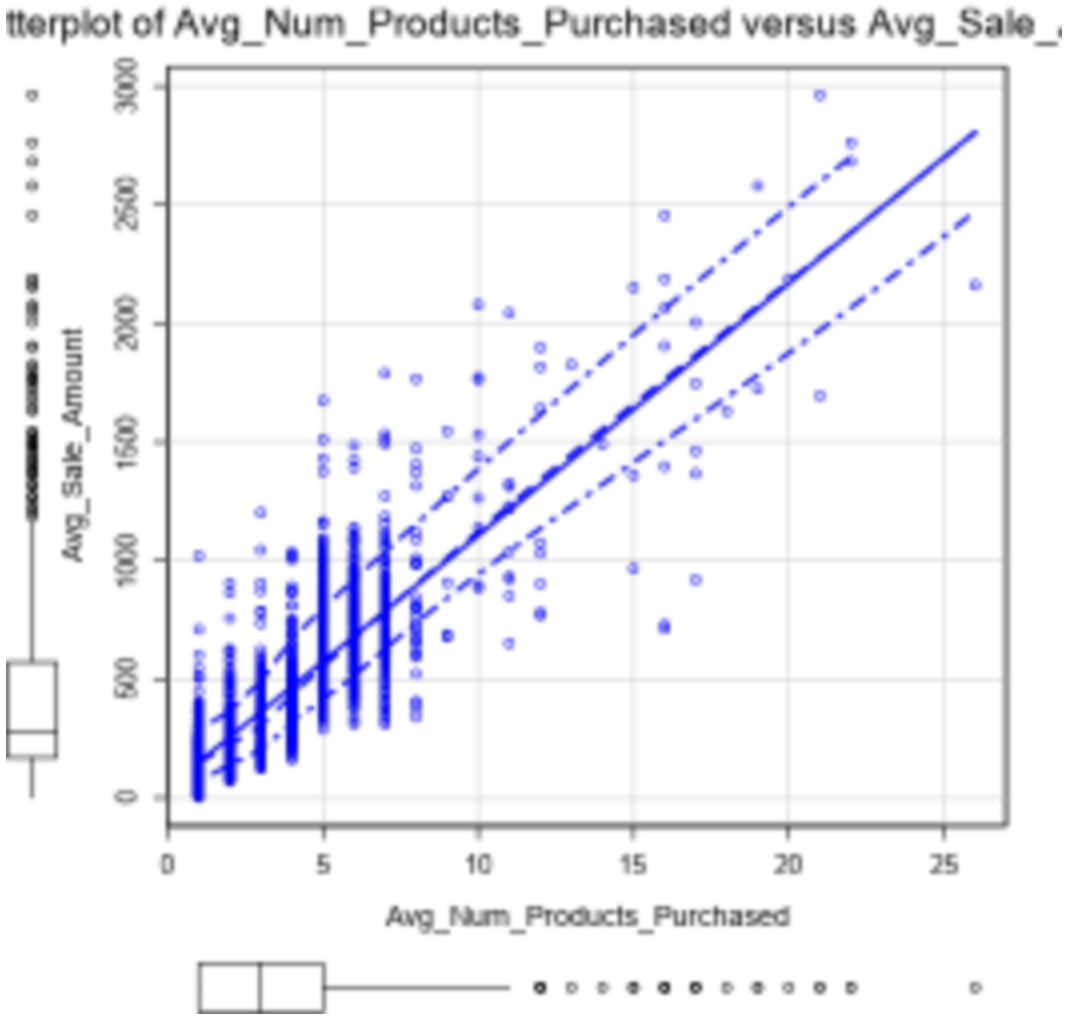
The dataset with the historical data of the 2,300 customers to build our model and the dataset with the 250 new customers from the mailing list in order to predict sales.

Step 2: Analysis, Modeling, and Validation

I used Alteryx and plugged in the “p1-customers.xlsx” dataset, then selected the linear regression model with “Avg_Sale_Amount” as the target variable and “Customer_Segment” and “Avg_Num_Products_Purchased” as the predictor variables.

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 *“Avg_Num_Products_Purchased”* variable after plotting all numerical variables’ scatterplots against the target variable *“Avg_Sale_Amount”.* As shown below there is a clear linear relationship between the two variables which makes it a good predictor. Later after running the model the p-value was also very small which shows that it is statistically significant.



Since the variable *"Customer_Segment"* is a categorical one and we cannot check linearity via the scatterplot, I tested it via trial and error along with other categorical variables and I found that the p-values were very small proving the statistical significance.

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.

As mentioned above the p-values of both predictor variables are very small:

- **Customer_SegmentLoyalty Club Only: < 2.2e-16**
- **Customer_SegmentLoyalty Club and Credit Card: < 2.2e-16**
- **Customer_SegmentStore Mailing List: < 2.2e-16**
- **Avg_Num_Products_Purchased: < 2.2e-16**

Also, the R squared values were quite high **(Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366)** which combined with the low P-values shows that the model is highly predictive.

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)

The best linear regression equation is as follows:

Revenue = 303.46 - 149.36*Customer_SegmentLoyalty Club Only + 281.84*Customer_SegmentLoyalty Club and Credit Card - 245.42*Customer_SegmentStore Mailing List + 0*Customer_SegmentCredit Card Only + 66.98*Avg_Num_Products_Purchased

Step 3: Presentation/Visualization

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

Yes, it should. The final prediction of expected profit is $21,987, more than double the extra profit the manager wanted in order to send the catalogs.

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

After building the model and training it with the *"p1-customers.xlsx"* dataset, I used the *"p1-mailinglist.xlsx"* to find the predicted revenue from the 250 new customers. Then I multiplied it with the probability of each one to buy the catalog and this was the expected revenue. The next step was to find the profit. To do so I multiplied the expected profit by the gross margin which is 50% and then subtract the $6.50 cost per catalog.

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

The expected profit is $21,987, as mentioned above.