Project 1: Predicting Catalog Demand

## **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?

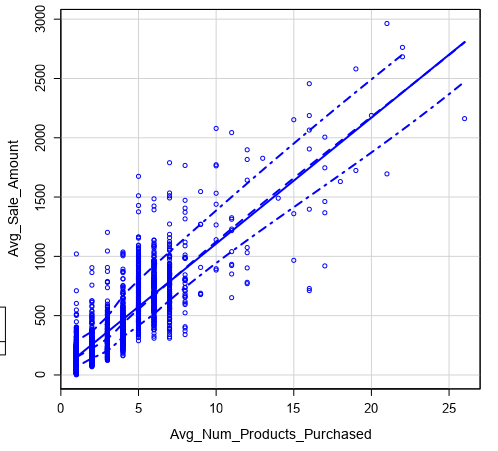
For this project, we need to decide whether it would be worth it to send catalogs out to new customers in order to bring in more business.

2. What data is needed to inform those decisions?

We will need to use the data from the previous year’s catalog to predict what our expected profit would be. We have data regarding the previous customers as well as data about the new customers.

## **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 used the average number of products purchased as my first variable as this will allow me to predict how much each new customer might purchase. The scatterplot above shows the relationship between the number of products purchased and the average sale amount.

I used the customer segment as a categorical prediction variable. Because this is not a numeric category, I used the linear regression tool in Alteryx to find the p-value of the column. P-values of less than .05 are significant and the result of this is very close to 0. Using the linear regression tool, we see value near 0 for the average number of products purchased also.

Because we are looking at multiple variables, we should also look at adjusted R-squared values. R-squared values closer to 0 mean that the model doesn’t explain any variations around the mean, whereas an R-squared value closer to 1 represents a model that explains all of the variations around the mean. With a R-squared value of .8366, we have a strong R-squared value, giving us a strong model.

The best linear regression equation based on this information is as follows:

Y = 303.46 – 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) + 66.98(Avg\_Num\_Products\_Purchased)

## **Step 3: Presentation/Visualization**

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

Upon completion of our analysis, I would recommend that the company send out the catalogs to the new customers. To come up with this recommendation, I used the Score tool in Alteryx to use our linear regression model to predict sales for these 250 customers. After getting a predicted sale amount, I took the percentage that each customer would make a purchase and multiplied that probability to each total.

After getting this predicted sale amount, I multiplied each of these totals by 50% to get the gross margin for each sale. Taking the overall total of all of this and subtracting the printing and distributing costs gives us a total expected profit of $21,987.51.