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?

* Using predictive analytics, we can estimate the potential profit from sending catalogs to the 250 new customers. Based on this estimate we can decide whether to send the catalogs or not.

1. What data is needed to inform those decisions?

* The data needed to inform the decisions is the data based on past sales and customer behavior information. Then we need to analyze data about customer base, purchase history and location, as well as information about the sales that occurred last year when the company sent out its first print. The probability that new customers will purchase items from our catalog, or not. Estimating sales for the current year and calculating expected profit, we can make an informed decision on whether to send the catalog or not.

## **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.***

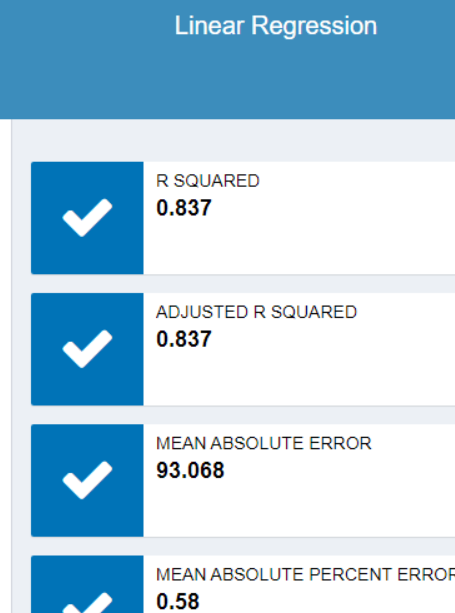
*At the minimum, answer these questions:*

1. How and why did you select the predictor variables in your model? 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 relationship of the available variables to the target variable (avg\_sale\_amount) was used to select predictor variables for the model.
* Scatterplots were used to determine the linear relationship between the numeric predictor variables and the target variable, and only the average number of products purchased and customer segments were found to be significant predictors of average sales. These are chosen as model predictor variables.

1. 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.

* The linear model is good because it only takes into account significant predictor variables as determined by the p-values and the adjusted r-square value of 0.837 is strong as well.



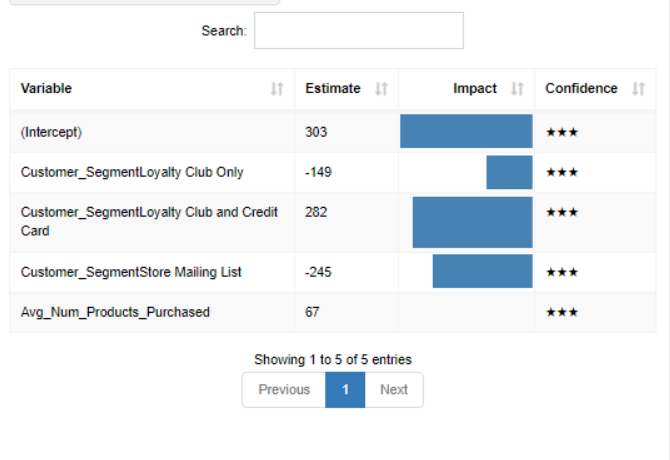
3. What is the best linear regression equation based on the available data? Each coefficient should have no more than 2 digits after the decimal (ex: 1.28)

**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.



Based on the avaialable data, the best linear regression equation will be:

Y = 303 – 149 \*(Customer\_segmentLoyalty Club only) + 282 \* (customer\_segmentLoyalty Club and credit card) -245 \* (customer\_SegmentStore Mailing List) + 67 \*(Avg\_num\_products\_purchased) + 0 \*(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, I will recommend that the company send the catalog to these 250 customers.

1. 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 this recommendation based on the final predicted profit of $22,389.55 if the catalog is sent, which is twice as much as the minimum requirement of $10,000. I mutiplied the score\_yes with the predicted score and while I got the sum of the expected profit revenue, and then I multiply by the percentage gross margin(50%), subracted the cost of catalog($6.50)of 250 customers. Formular = 48029.1 \* 50% - $6.50(250)

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

* $22,389.55