

**Project 1: Predicting Catalog Demand**  
**By : Neavil Porus A**

**Step 1: Business and Data Understanding**

**Key Decisions:**

**1. What decisions needs to be made?**

- Is it worthful enough to deliver the catalogs to new clients?

**2. What data is needed to inform those decisions?**

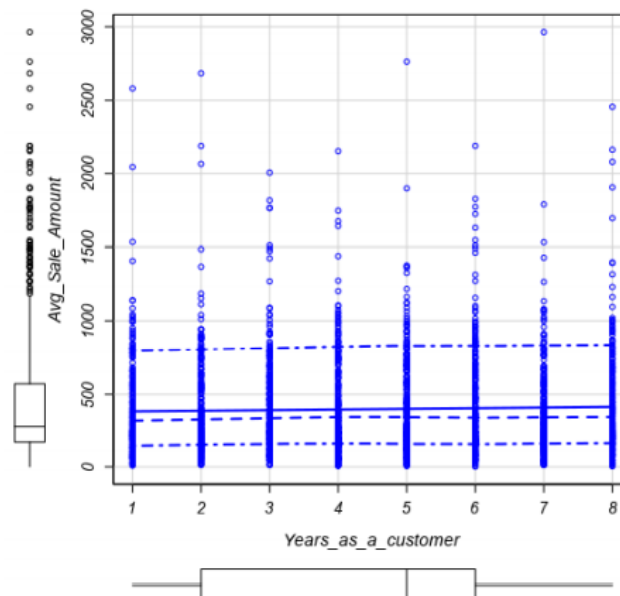
- We need data about the current customers and the new customers. Information based on the customer segment data and average numbers product purchased will help us to inform those decisions.

## **Step 2: Analysis, Modeling, and Validation**

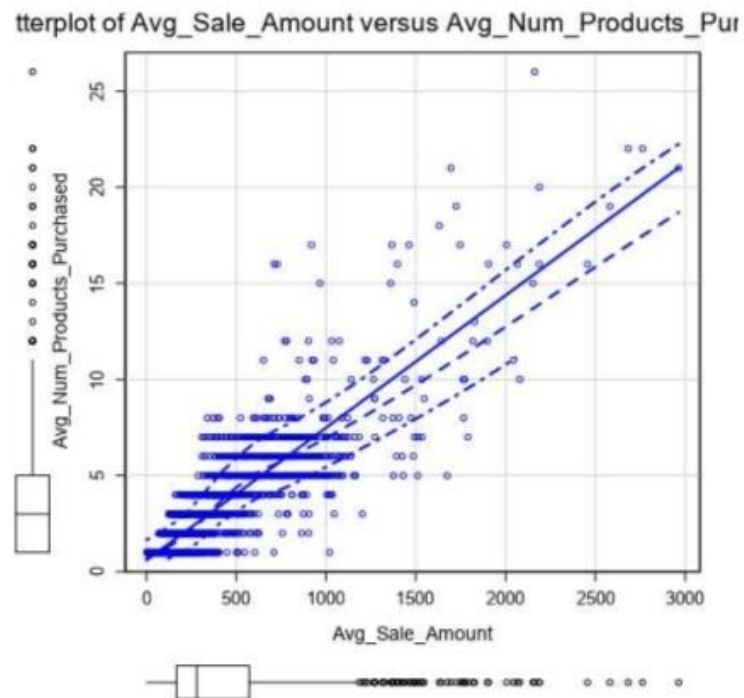
*Important: Use the p1-customers.xlsx to train your linear model.*

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 have used **scatterplots** and **linear regression tool** in Alteryx to find an answer for this one. Below is **one example showing no correlation** – which is I didn't used it in my model.



The **only numerical value** that can be useful during my project is **Avg\_number\_of\_products** purchased so this data can be used for **improvising our model**.



From the table below we can see what the **P-value for various data available** in the data set. The **customer segment and Avg\_num\_Products\_purchased** P-value is less than 0.05 which can be used for our model analysis.

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.

✓	R SQUARED 0.837
✓	ADJUSTED R SQUARED 0.837
✓	MEAN ABSOLUTE ERROR 93.068
✓	MEAN ABSOLUTE PERCENT ERROR 0.58
✓	MEAN SQUARED ERROR 18861.84
✓	ROOT MEAN SQUARED ERROR 137.338
✓	F-STATISTIC 3039.74 on 4 and 2370 degrees of freedom
✓	RESIDUAL STANDARD ERROR 137.483 on 2373 degrees of freedom

In the table from the **Report from linear regression** we see both **P-value and Statistical Significance**. I have used only data that is **Statistically significance and with the small P-value**.

We go R Squared = 0.837, Adjusted R Squared = 0.837.
--

At the same time, the **multiple R square also has a high value** [ $R^2 > 0$  &&  $R^2 > 0.8$ ], so I don't have concerns regarding the **quality of the 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

This is the best regression equation that we could create using this data:

**Y = Intercept + b1 \* Variable\_1 + b2 \* Variable\_2 + b3 \* Variable\_3**

*Y = 303.46 + b1 \* (Loyalty Club Only) + b2 \* (Loyalty Club and Credit Card) + b3 \* (Store Mailing List) + b4 \* (Avg.Num.Products.Purchased) [Formula]*

**Y = 303.46 + -149.36\*(Loyalty Club Only) + 281.84 \*(Loyalty Club and Credit Card) + (-245.42)\*(Store Mailing List) + 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, the company should deliver the catalogs to new customers since this will bring **more than 10k** of profit for the company.

2. How did you come up with your recommendation?

I have used a **linear regression model** to provide my recommendation. I have enriched the data set with prediction information which can be used for predicting future sales.

Using **linear regression model**, the **expected revenue** from each customer is **determined by multiplying** expected sale amount with **Score\_Yes** value.

With a **gross margin of 50%, 50%** is **deducted from the sum of expected revenue** before the **cost of catalog (\$6.50)** is **subtracted to obtain total predicted revenue** for **250 customers**.

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

The expected profit is equal to: **21987.44**.

