

# Predicting Catalog Demand

Atheer Hassan Alzahrani  
atheeralzh@gmail.com

## Step 1: Business and Data Understanding

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

The decision is predicting how much money the company can expect to earn from sending out a catalog to 250 new customers.

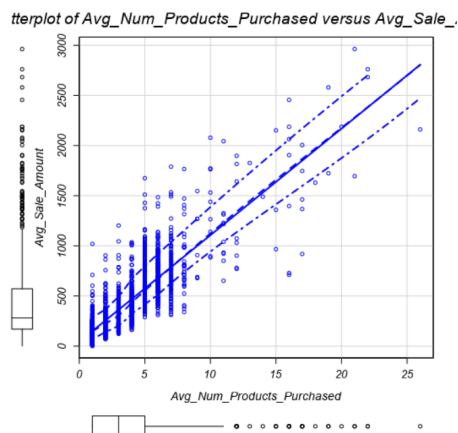
### 2. What data is needed to inform those decisions?

We need customers data, mailing data in Excel sheet  
and two variables: Numerical variable and Categorical variable i.e. Avg\_Num\_Products\_Purchased, Customer Segment.  
Cost of Catalogue \$6.50 and gross margin 50%.

## Step 2: Analysis, Modeling, and Validation

### 3. How and why did you select the predictor variables in your model?

For this model, The Avg\_Sale\_Amount is the target variable and that only two predictor variables are statistically significant Customer\_Segments and Avg\_Num\_Products\_Purchased so we can leave them in because these have P-value less than 0.05. Let's take a look at the adjusted R-squared. It's above 0.8, which is good because The higher the r-squared, the higher the explanatory power of the model.



From the Scatter plot would indicate the (Avg\_Num\_Products\_Purchased) would be a good candidate to be a predictor variable for the target variable (Avg\_Sale\_Amount), because have a positive relationship.

For categorical variables, I have included a screenshot the Linear Regression Report below and We are looking for P-values less than 0.05, So would be a good candidate to be a predictor variable for the target variable (Avg\_Sale\_Amount).

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.

| 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   |             |            |         |               |
| Response: Avg_Sale_Amount  |             |            |         |               |
|  | Sum Sq      | DF         | F value | Pr(>F)        |
| Customer_Segment   | 28715078.96 | 3          | 506.4   | < 2.2e-16 *** |
| Avg_Num_Products_Purchased   | 36939582.5  | 1          | 1954.31 | < 2.2e-16 *** |
| Residuals  | 44796869.07 | 2370       |         |               |

From the table above two variables are statistically significant Customer\_Segments and Avg\_Num\_Products\_Purchased because these have P-value less than 0.05.

Let's take a look at the adjusted R-squared value 0.8366, which is good. So this good model because the p-value and R-squared value are statically significant.

2. What is the best linear regression equation based on the available data?  

$$\text{Avg\_Sale\_Amount} = 303.46 - 149.36 * (\text{customer\_segment: loyalty club only}) + 281.84 * (\text{customer\_segment: loyalty club \& credit card}) - 245.42 * (\text{customer\_segment: Mailing list})$$

### Step 3: Presentation/Visualization

1. What is your recommendation? Should the company send the catalog to these 250 customers?  
 Yes, the company should send catalogues to these 250 customer.
2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)  
 First calculate the Avg\_Sales using the linear regression model. Then multiplied the Score by Score\_yes. Then calculate the profit using 50% margin and \$6.50 cost of each catalogue and for 250 customer.
3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?  

$$\text{Profit} = (\text{profit} * .5) - (6.50 * 250)$$

$$= 21,987.43587$$

- Alteryx workflow

