<u>Project 1: Predicting Catalog Demand</u>

## Business and Data Understanding

**Key Decisions:**

1.  The main decision here is company wants to know that if the expected profit from these new customers will be minimum $10,000 or not then will decide to send the catalogue or not to the customers.

2. To calculate the expected profit from customers.

   **Data required:**

Customer data :
1. Average sale amount
2. Average no. of products purchased
3. Customer segment
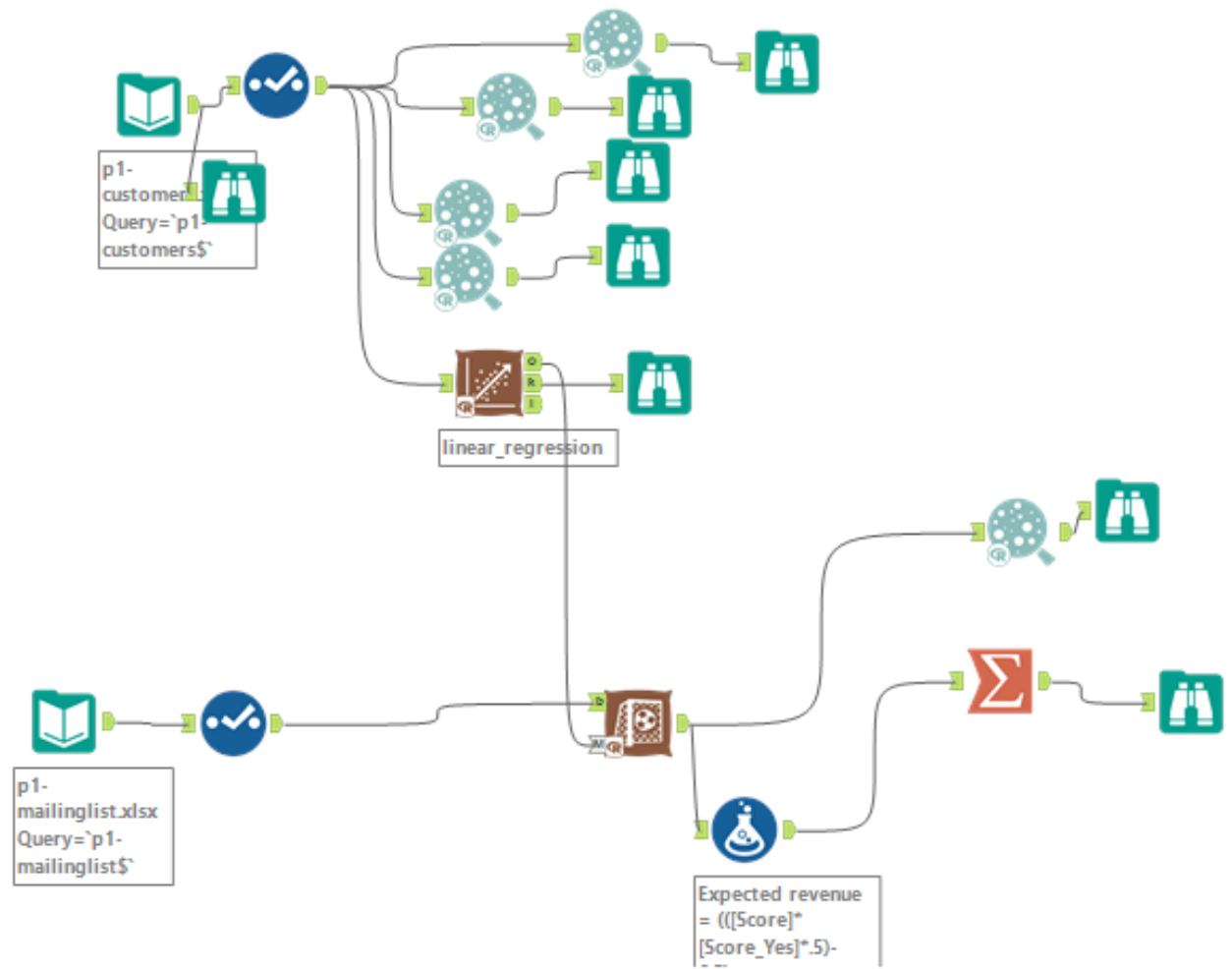4. Score yes

## Analysis, Modeling, and Validation

To select the predictor variables we need to plot a scatterplot between target variable and predictor variables here only variable**: Average no. of products purchased** is showing the linear relationship with the **target variable : Average Sale amount**. That signifies it is the good predictor variable and clearly impacting the sales and profit.

tterplot of Avg_Num_Products_Purchased versus Avg_Sale_
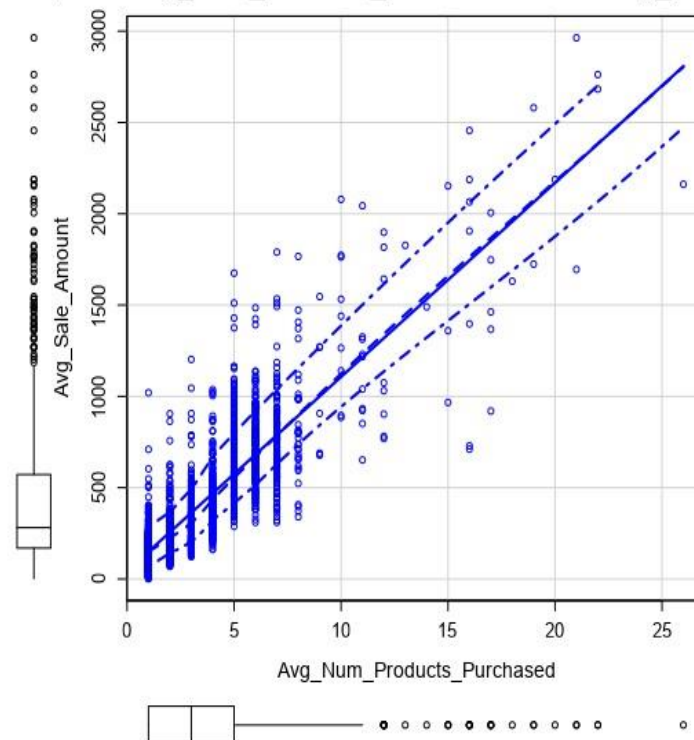


Second predictor variable used in the linear regression model is the customer_segment  has p-value less than 0.05 that shows it is statistically significant for the model.

This linear regression model is good because here the R-squared value is 0.83 that is consider to be good for the linear regression model and adjusted R-squared is also same which clearly shows the strength of model.

p1-
customer
Query=`p1-
customers$`

linear_regression

p1-
mailinglist.xlsx
Query=`p1-
mailinglist$`

Expected revenue
= (([Score]*
[Score_Yes]*.5)-

1.

| Record | Report |
|---|---|

**Report for Linear Model linear_regression**

*Basic Summary*

Call:

lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)

Residuals:

| Min | 1Q | Median | 3Q | M |
|---|---|---|---|---|
| -663.8 | -67.3 | -1.9 | 70.7 | 971 |

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 | | | |

Both the target variables that are customer segment and Average no. OF products purchased has p-value less than 0.05 which is statistically significant and indicate the strong model.

**Linear regression equation:** Y = 303.46+ 66.98* Average No. of products purchased – 149.36 * Customer segment Loyalty club only + 281.84 (If Type: Credit Card and loyalty club) – 245.42 (If Type: Store mailing list) + 0 (If Type: Cash)

# Presentation/Visualization

1.  Yes, company should send the catalogues to these 250 new customers as the expected profit is good and within the range of that the company expected. Sending the catalogues to these new customer will increase the sales of company and so the profit.

2. Firstly we calculated average sales by using linear regression model.
Multiply the result with score_yes and 0.5 (profit margin is 50%)
Then, subtract the result from the cost of printing and distributing of 250 catalogues ($6.5*250)
Final result = Sum(average sales * score yes *0.5) – (6.5 * 250)

On analyzing the dataset of customers it shows that two factors that are customer segment and average no of products purchased are affecting the Average sale amount of company And are also statistically significant on that basis and including these variable In the model it is showing the expected profit greater than $10000 and indicating that sending catalogue to these new 250 customers would increase the sales of company and bring large profit.

3 .Total expected profit for 250 customers is: 21,987.4356