Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Two datasets Mailing list and customer are provided to us. The company has 250 new customers from their mailing list that they want to send the catalog to. Management does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000.

Decisions to be made:-
1. How much profit the company will get by adding 250 customers?
2. Should they send the catalog to all the customers?

The information necessary to make these decisions are:-
1. Predictor variable
   a. The Customer segment which is categorical variable is a predictor variable.
   b. Average_number_products_purchased is also important predictor variable.
2. The Score_Yes variable tells us whether the customer will buy the product or not.
3. Target variable
   a. Average sales.
      Average sales needs to be predicted first and then gross margin and net profit.
4. The cost of making and transporting the catalog should also be known i.e. $6.50

Step 2: Analysis, Modelling, and Validation

1. The linear regression model:-

The linear regression model and the workflow is designed in Alteryx. As mentioned above, we have taken two predictor variables for the target variable.

1. On considering the Avg_Num_Products_Purchased , the p value is $2.2*e^{16}$.

:rplot of Avg_Num_Products_Purchased versus Avg_Sale
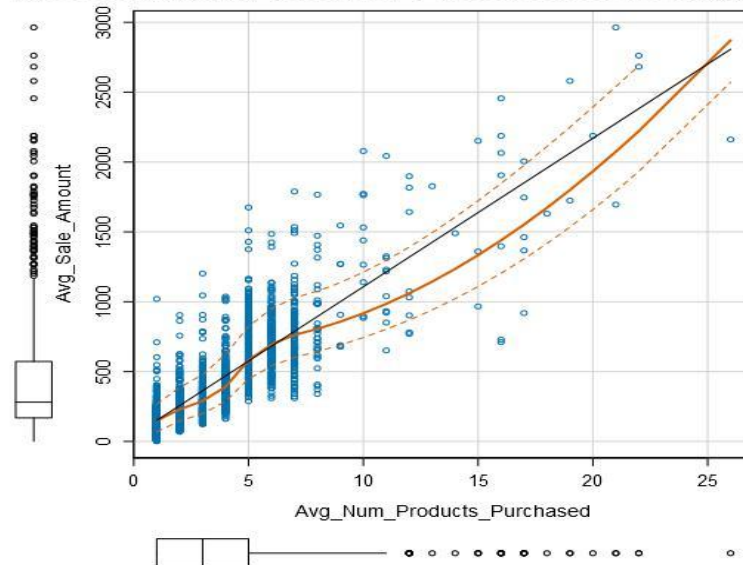
2. The second target variable i.e Catalog segment is also significantly important as it has its p value as $2.2*e^{16}$.

3. The value of Adjusted r squared is 0.8366 which indicates that our model is pretty good .

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

From the above report, the linear regression equation formed is

Y = 303.46 -149.36 (Customer_SegmentLoyalty Club Only) + 281.84 (Customer_SegmentLoyalty Club and Credit Card) – 245.42 (Customer_SegmentStore Mailing List) + 0*(Customer segment_credit_card only) +66.98 (Avg_Num_Products_Purchased)

Step 3: Presentation/Visualization
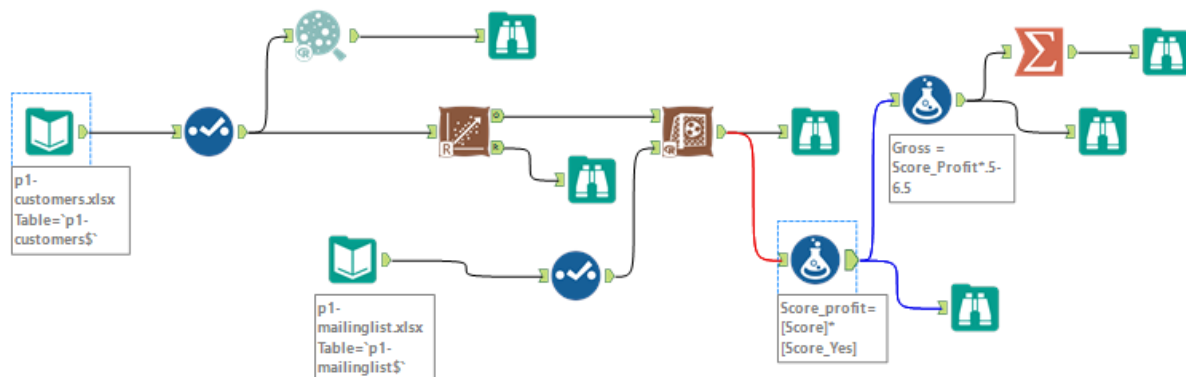
This workflow has been created in the Alteryx.
I have reached at the following conclusions by using this workflow and other resources available:-

1. Catalog should be delivered to customers by company as the net profit > $ 10000 .

   I reached this conclusion by the following steps :-

   a. Average sales amount of a customer is calculated using the regression equation.
   b. Then, the amount is multiplied by variable Score_Yes i.e. the probability of a customer buying a product.
   c. After that, the total profit is halved (50%) and gross margin is obtained .
   d. Later, the amount of production of a catalog i.e. $6.50 is subtracted from the gross margin.
   e. Finally, Summation of profit for all the customers is done , hence providing us with the net profit.

❖ The final expected output is **$21987.43**