Predicting Catalog Demand

Project Overview

Last year, the company sent out its first print catalog, and is preparing to send out this year's catalog in the coming months. The company has 250 new customers from their mailing list that they want to send the catalog to. It needs to be determined how much profit the company can expect from sending a catalog to these customers. Management does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000.

Step 1: Business and Data Understanding

1. What decisions needs to be made?
   The decision needs to be made is whether to send the catalogs to the new customers on the mailing list. The company will only send the catalog if the profits exceed $10,000

2. What data is needed to inform those decisions?
   Data needed to inform decisions are average number of products purchased, average sales amount, customer segment, margin and cost of catalog.

Step 2: Analysis, Modeling, and Validation

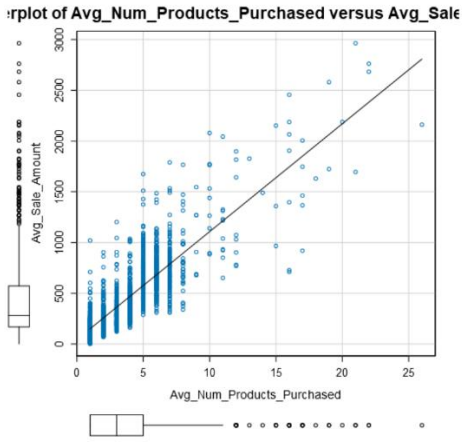

Figure 1. Plot of Average number of Product Purchased vs Average Sales Amount

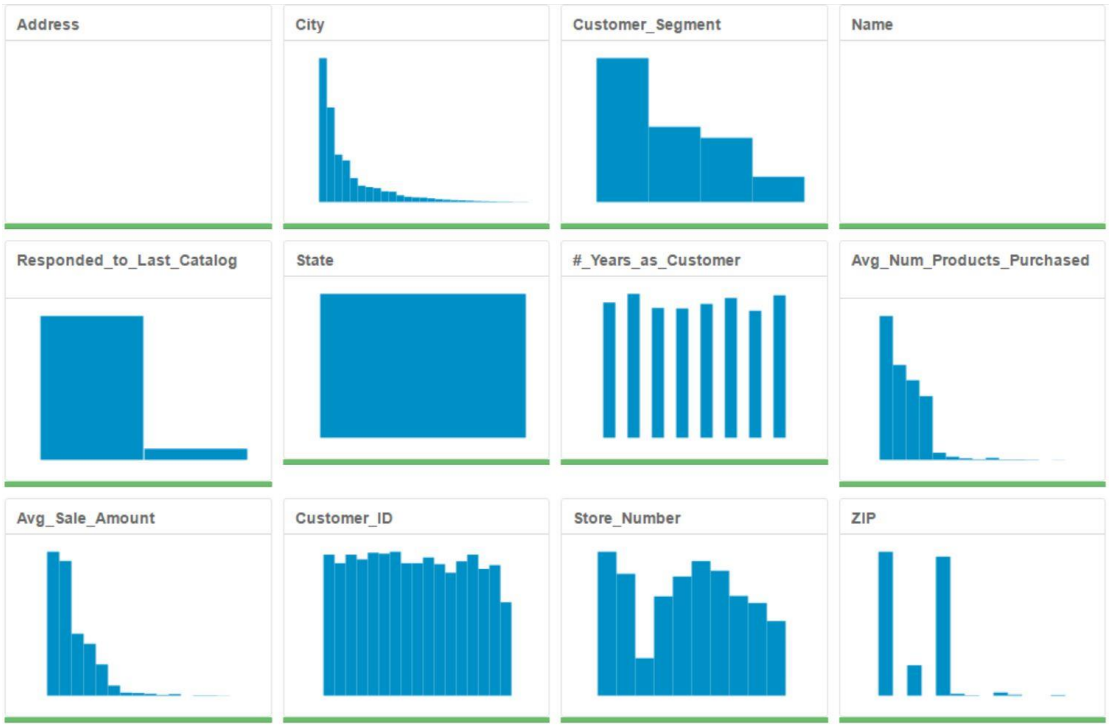

Figure 2. Field Summary of All Variables

Record | Report

Report for Linear Model Linear_Regression_Customer

Basic Summary

Call:
lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = inputs$the.data)

Residuals:

Min	1Q	Median	3Q	Max
-663.8	-67.3	-1.9	70.7	971.7

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 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			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 3. Linear Regression report

According to figure 1, Ave_num_products_purchased is selected as variable, because it has linear relationship with the Average_sales_amount.

Figure 2 indicates that he customer segment (p-value <0.05) is another variable which has influence to the Average_sales_amount.

Other variables don't present obvious linear relationship with the Average_sales_amount, based on figure 2.

Figure 2 contains the information which shows the model is a good predict. The P-value for each of the four variables are <2.2e-16 and follow with *** means the relationship between target variable and predictors are statistically significant. In this model, the adjusted R-square is 0.8366, which is more than 0.7, means the explanatory power of the model is strong.

Thus the linear regression equation is :
Y = 303.46 -149.36* Customer_SegmentLoyalty Club Only + 281.84 * Customer_SegmentLoyalty Club and Credit Card – 245.42* Customer_SegmentStore 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?
The company should send the catalog to the 250 new customers.

2. How did you come up with your recommendation?
After scoring the new customers list with the linear regression model, the average gross margin (price - cost) on all products sold through the catalog is 50%, after the revenue*gross margin*50%, the cost of the printing and distributing per catalog of $6.5 need to be subtracted.

3. What is the expected profit from the new catalog
The expected profits calculated by Alteryx is $21987.43, which exceed the amount of $10,000

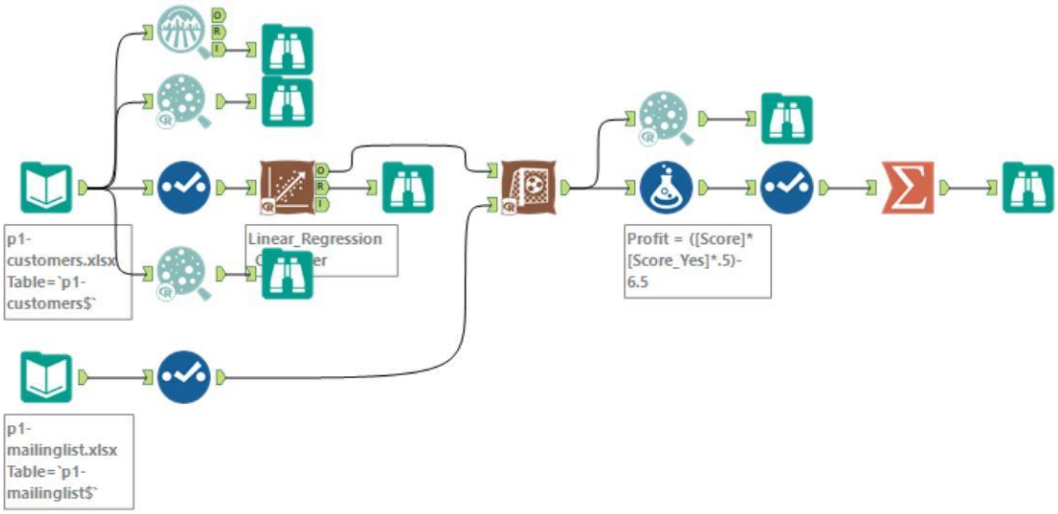

Figure 4. Workflow