

Predicting Catalog Demand

Title - Predicting Catalog Demand

Description - Predicting whether to send a new catalogs coming next month to new customers or not

Problem Statement - Companies need to make decisions about sending new catalog to customers based on specified conditions.

Approach and Technique Used - Multiple regression models perfectly work for this problem.

Tool - Alteryx

What I Learned - Build multiple linear regression model to predict real value
Training multiple linear regression model
Evaluate multiple linear regression model
Make prediction using the trained model

This project is part of the **Predictive Analytics for Business** NanoDegree program offered by Udacity.

Step 1: Business and Data Understanding

1. What decisions need to be made?

The company manufactures and sells high-end goods. They need to make the decision about whether to send a new catalog coming next month to new customers or not.

2 . What data is needed to inform those decisions?

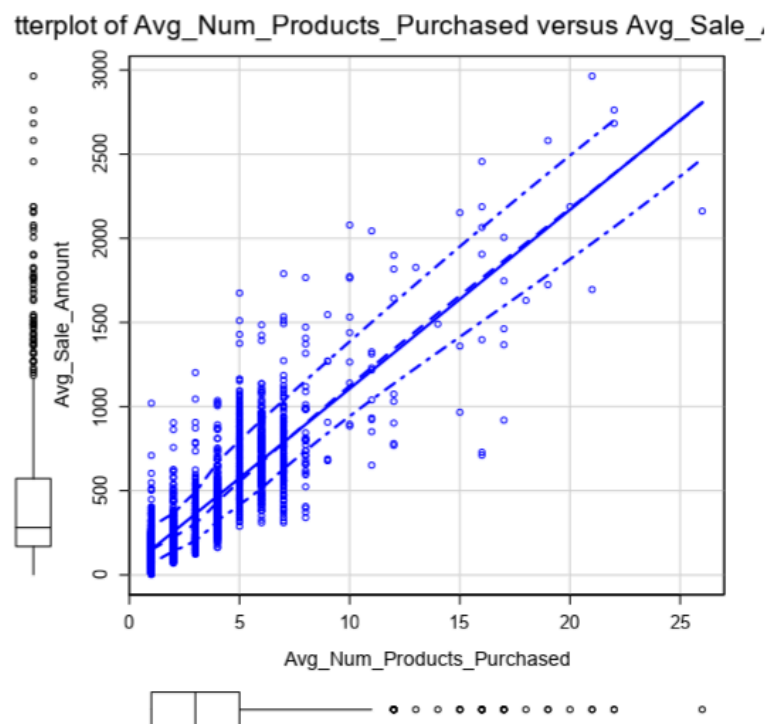
We need a dataset with features that are important to solve this problem such as Avg_sale_Amount, Customer_Segment, Avg_Num_Products_purchased, Score_no, Score_Yes, gross margin, and cost.

Step 2: Analysis, Modeling, and Validation

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

I selected a number of numeric predictor variables and created a scatter plot using Alteryx to check the relationship between the target variables. But the only numeric predictor variable that has an expected correlation, is Avg_Num_Products_Purchased. The Scatter plot is listed below.

1



I selected the predictor variable as Customer_Segment because it is statistically significant to our target variable. It has a higher correlation with target than other variables. I listed the chart below.

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.379e+03	2.149e+03	-0.6416	0.52118
Customer_SegmentLoyalty Club Only	-1.497e+02	8.980e+00	-16.6659	< 2.2e-16 ***
Customer_SegmentLoyalty Club and Credit Card	2.824e+02	1.193e+01	23.6659	< 2.2e-16 ***
Customer_SegmentStore Mailing List	-2.459e+02	9.774e+00	-25.1627	< 2.2e-16 ***
Customer_ID	-1.373e-03	2.941e-03	-0.4669	0.64063
ZIP	2.248e-02	2.660e-02	0.8451	0.39814
Store_Number	-1.011e+00	1.007e+00	-1.0042	0.31539
Avg_Num_Products_Purchased	6.700e+01	1.517e+00	44.1582	< 2.2e-16 ***
X_Years_as_Customer	-2.345e+00	1.223e+00	-1.9167	0.0554 .

2. Explain why you believe your linear model is a good model?

I selected the numeric predictor variable as Avg_Num_Products_Purchased. It has a p-value below 0.05 so statistically significant with our target Variable. It has a strong correlation with the target variable. Another variable is a categorical variable. It is Customer_Segment. It also has a p-value of 0.05. It has a strong correlation with our target variable. At the same time, the R-square value is 0.8369 and the adjusted R-squared is 0.8366 so I am very confident with this model.

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

3. What is the multiple linear regression equation based on the available data?

$$Y = 303.46 + -149.36 * (\text{If Customer_SegmentLoyalty Club Only}) + 281.84 * (\text{If Customer_SegmentLoyalty Club and Credit Card}) + -245.42 * (\text{If Customer_SegmentStore Mailing List}) + 66.98 * \text{Avg_Num_Products_Purchased}$$

Step 3: Presentation/Visualization

1. What is your recommendation? The company should send the catalog to these 250 customers who have exceeded the threshold.

2. How did you come up with your recommendation?

Expected Profit = (Total revenue * Gross margin) - (catalog cost * number of customers)

$$\text{Expected Profit} = 47224.87 * 0.5 - (6.5 * 250)$$

$$\text{Expected Profit} = 23612.43 - 1625$$

$$\text{Expected Profit} = 21987.43 \text{ US Dollars.}$$

Profit is greater than 10 000 US Dollars, so the company should send the catalog to those 250 customers.

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

Expected Profit = 21987.43 US Dollars.

Alteryx WorkFlow

