

Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

We have to analyze the data and see if sending catalogs to new customers will be beneficial for the company or not.

2. What data is needed to inform those decisions?

We need to build linear regression model on historical data. Apply that model to predict average sale amount from 250 new customers. The cost for printing and distribution is \$6.50 and average gross margin from sold product through catalog is 50%. So, expected profit = SUM (avg. sale amount * score_yes) * 0.5 – 6.5*250. If expected profit is >=10000 then company should send catalog.

Step 2: Analysis, Modeling, and Validation

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

By exploring the data, I knew that state has the same value and there is no case that two different customers has same address. Name and customer_id do not affect the sales amount so I did not select that. Response column was useless to use in the model as that is not going to use in mailing list.

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

Record

1

Report

Report for Linear Model Linear_model_for_catalog_project

2

Basic Summary

3

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

4

Residuals:

Min

-663.8

1Q

-67.3

Median

-1.9

3Q

70.7

Max

971.7

6

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 ***

8

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

8

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

9

Type II ANOVA Analysis

10

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

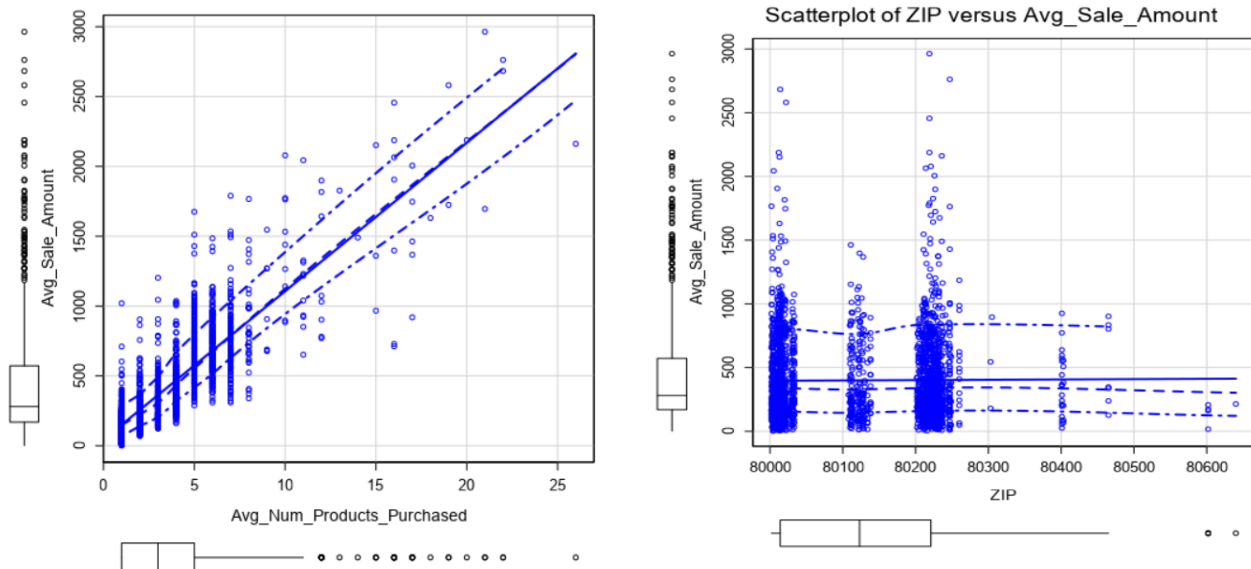
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

This is the report based on created linear regression model. As we can see, predictor variables have p-value below 0.05, relation between target variable and avg_sale_amount is significant. R-square and adjusted R-squared value is 0.8369 and 0.8366, which is very high. Considering all this the linear regression model is good for this scenario.

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

Avg_Sale_Amount = 303.46 + 66.98 * Avg_Num_Products_Purchased -149.36 (If Customer_Segment: Loyalty Club Only) + 281.84 (If Customer_Segment is Loyalty Club and Credit Card) – 245.42 (If Customer_Segment is Store Mailing List) + 0 (If Customer_Segment is Credit Card Only)

Step 3: Presentation/Visualization



1. What is your recommendation? Should the company send the catalog to these 250 customers?

My recommendation is company should send catalog to all 250 new customers. Because the profit from all the new customers is \$21987.44 which is >10000.

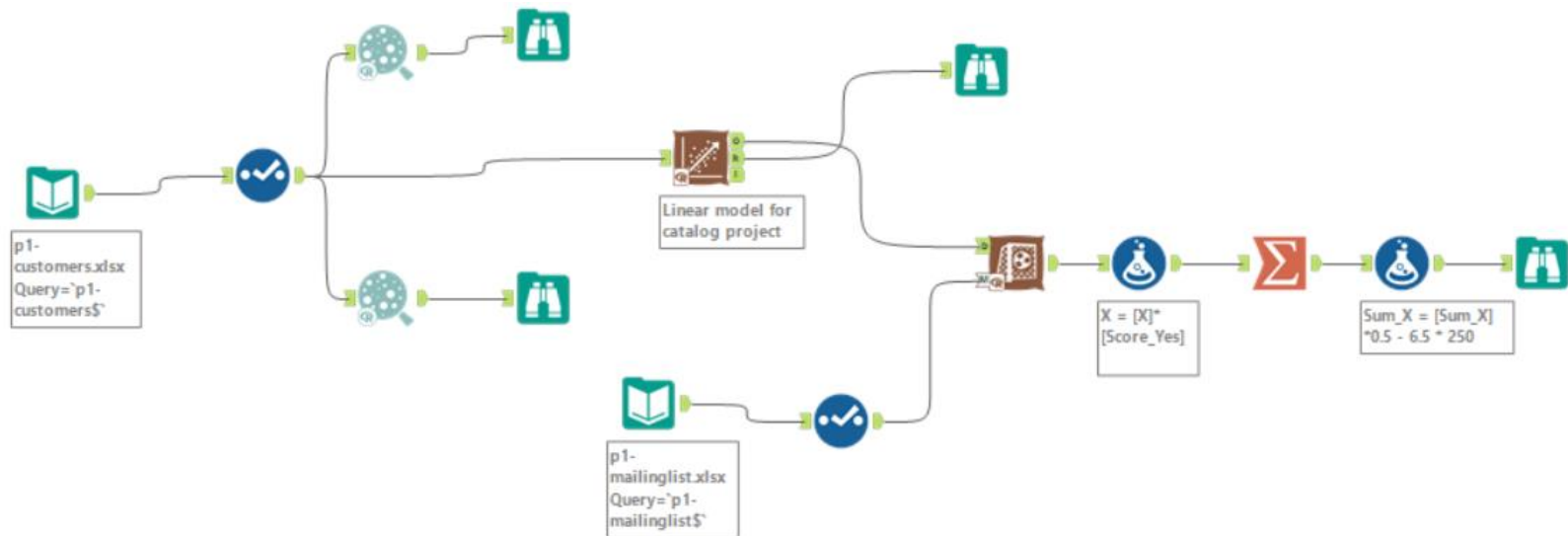
2. How did you come up with your recommendation?

First, I created linear regression model from the historical data. From that model I created formula. That formula is,

Avg_Sale_Amount = 303.46 + 66.98 * Avg_Num_Products_Purchased -149.36 (If Customer_Segment: Loyalty Club Only) + 281.84 (If Customer_Segment is Loyalty Club and Credit Card) – 245.42 (If Customer_Segment is Store Mailing List) + 0 (If Customer_Segment is Credit Card Only)

Loyalty Club and Credit Card) – 245.42 (If Customer_Segment is Store Mailing List) + 0 (If Customer_Segment is Credit Card Only)

Then I applied this to mailing list dataset to get predicted value of avg_sale_amount. I applied all these things by Alteryx. The image below shows the calculations and model built in Alteryx.



3. What is the expected profit from the new catalog?

The image above shows the model I built in Alteryx. The model gives the output of expected profit of \$21987.43.

Record	Sum_X
1	21987.435687