

Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions that needs to be made:

1. What decisions needs to be made?
 - Is it worth to send the catalogs to the new clients?
2. What data is needed to inform those decisions?
 - We need data about the current customers and the new customers. Information about the customer segment and average numbers product purchased will help us with answering this question.

Step 2: Analysis, Modeling, and Validation

1. How and why did you select the predictor variables in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer back to the "Multiple Linear Regression with Excel" lesson to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.
 - I have used scatterplots and linear regression tool in Alteryx to find an answer for this one. Below is one example showing no correlation – which is I didn't used it in my model.

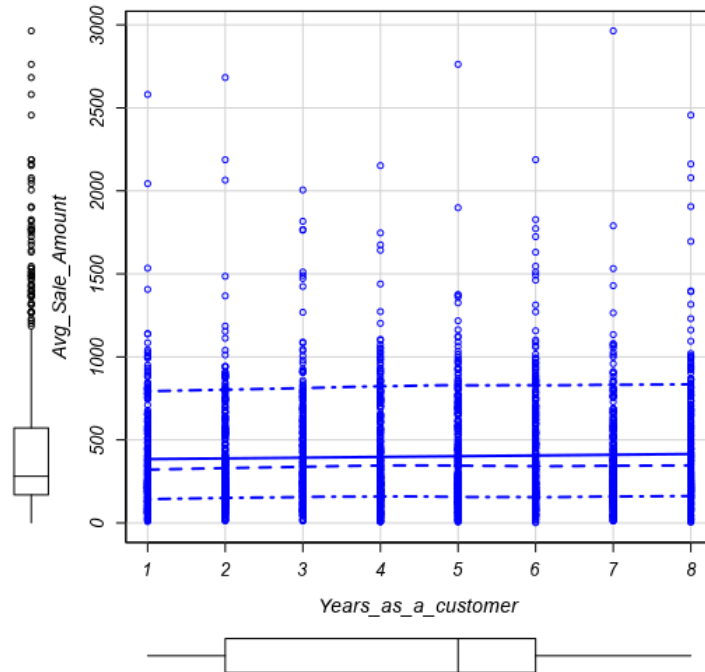


Figure 1. Scatterplot showing Avg Sale Amount VS Years as a Customer

The only numerical value that can be useful during my project is avg number of products purchased, and I will use it in my analysis.

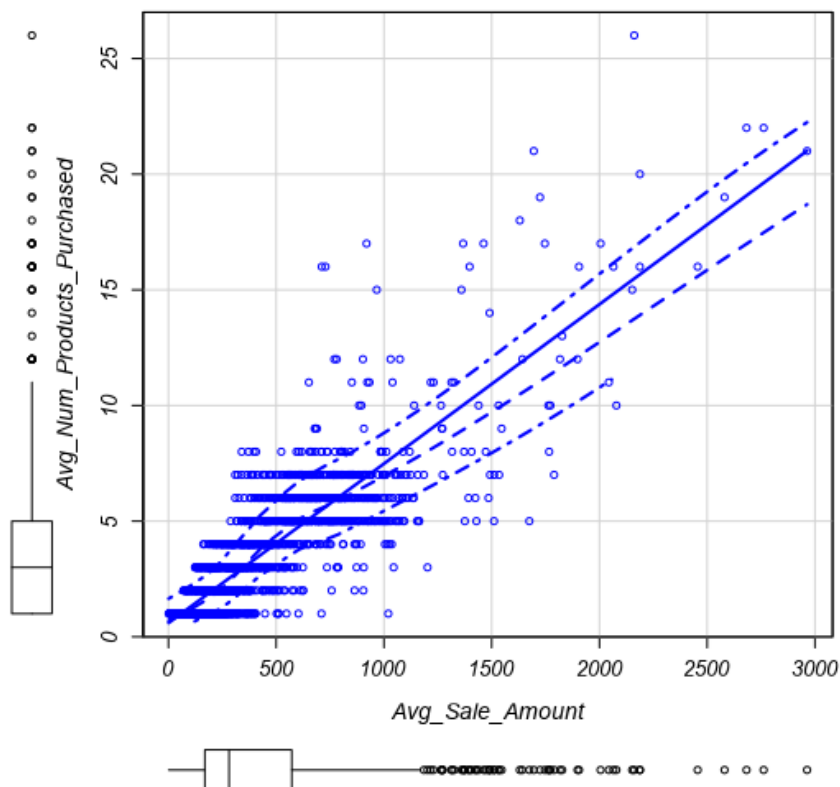


Figure 2. Scatterplot showing Avg Number of Products Purchased VS Avg Sale Amount

From the table below we can see what the P-value for various data available in the data set. The customer segment and AVG_num_Products_purchased P-value is less than 0.05, so I have decided to use them in the analysis.

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1384.1983	2.149e+03	-0.6441	0.51958
Customer_SegmentLoyalty Club Only	-149.5782	8.977e+00	-16.6625	< 2.2e-16 ***
Customer_SegmentLoyalty Club and Credit Card	282.6768	1.191e+01	23.7335	< 2.2e-16 ***
Customer_SegmentStore Mailing List	-245.8485	9.770e+00	-25.1625	< 2.2e-16 ***
ZIP	0.0225	2.659e-02	0.8460	0.39761
Store_Number	-1.0002	1.006e+00	-0.9939	0.32037
Avg_Num_Products_Purchased	66.9646	1.515e+00	44.1928	< 2.2e-16 ***
Year	-2.3528	1.223e+00	-1.9239	0.05449 .

Figure 3. Report for Linear Regression Model

2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that your model produced.

In the table from the Report from linear regression model (figure 3), we see both P-value and Statistical Significance. I have used only data that is Statistically significant and with the small P-value.

At the same time, the multiple R square also has a high value, so I don't have concerns regarding the quality of the model.

Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366

Figure 4. Value from: Report for Linear Regression Model

3. What is the best linear regression equation based on the available data? Each coefficient should have no more than 2 digits after the decimal (ex: 1.28)

This is the best regression equation that we could create using this data:

$$\text{Avg_Sale_Amount} = 303.46 - 149.36 * (\text{Customer_Segment : Loyalty Club Only}) + 281.84 * (\text{Customer Segment : Loyalty Club and Credit Card}) - 245.52 * (\text{Customer Segment : Store Mailing List}) + 66.984 * (\text{Avg_Num_Products_Purchased})$$

Step 3: Presentation/Visualization

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

Yes, the company should send the catalogue to the new customers as well as this will bring more than 10k of profit for the company.

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)

I have used a linear regression model to provide my recommendation. I have enriched the data set with prediction information, and I have used it as a predictor of future sales.

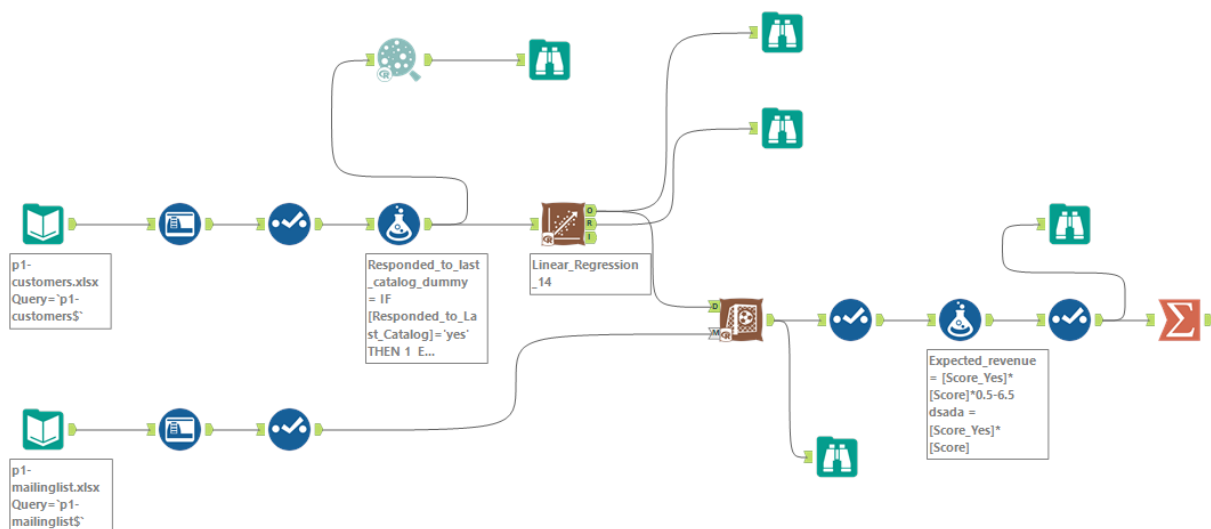
From each position, I subtracted the cost of the magazine (6,5).

Thanks to that I have calculated expected profit.

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

The expected profit is equal to: 21987,44.

Alteryx Workflow:



Workflow 1. Predicting the Sales Using Linear Regression Model