

Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

1. What decisions needs to be made?

Whether the company should send the catalogs to the 250 new costumers and what is the expected profit.

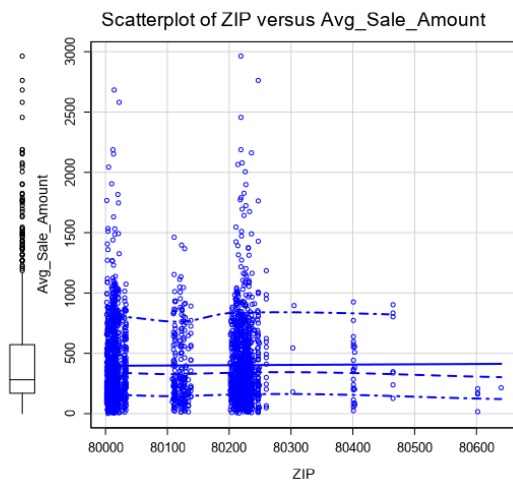
2. What data is needed to inform those decisions?

- Last year sales data set
- Predict average number of sales.
- Predict the expected revenue.
- Printing cost = 6.5
- Gross margin %50
- Expected profit ≥ 10000
- Linear regression model.

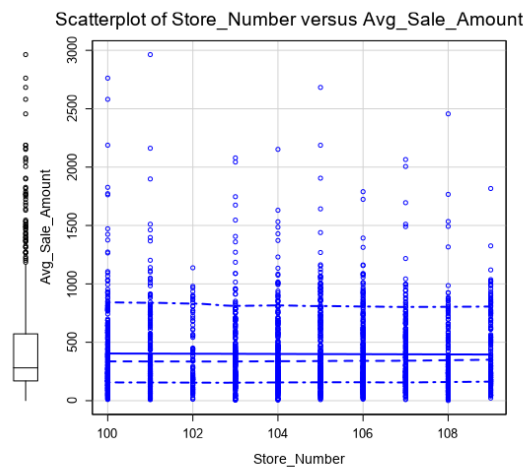
Step 2: Analysis, Modeling, and Validation

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.

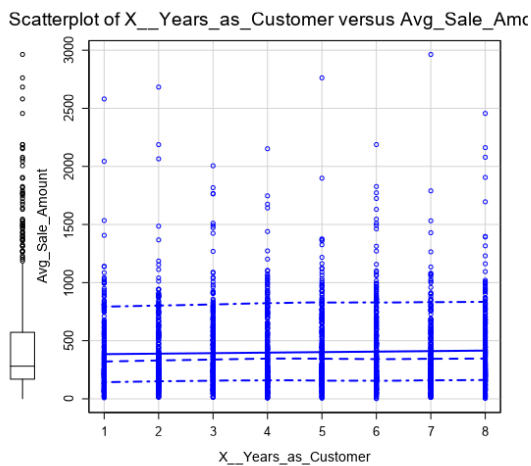
The target variable is Avg_Sale_Amount. The predictor variables were chosen based on the scatterplots of average sale amount versus the potential predictor variables for better understanding of the data correlations. But first by just looking at the data, Name, Customer_ID, Adress, State, and Responded_to_Last_Catalog will not be selected as they will not be beneficial to the model.



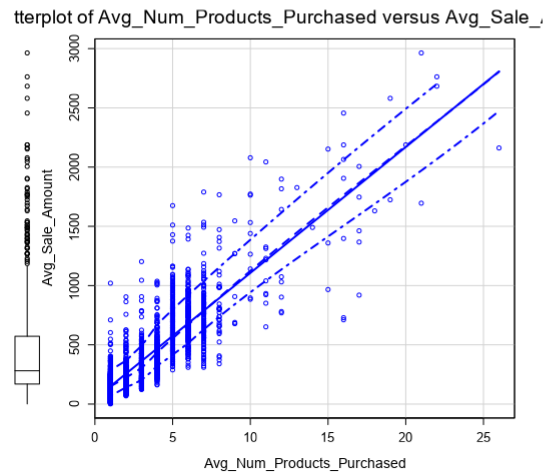
Not linear



Not linear



Not linear



Weak linear relationship

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.

Our model consider to be a reliable model since it has a low P-value and a high R-squared value with a P-value= $<2.2e-16$ and R-squared=0.8366.

Record

Report

1

Report for Linear Model Linear_Regression

2

Basic Summary

3

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

4

Residuals:

5

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

6

Coefficients:

7

	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

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

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

$$Y = 303 + (-149.36 * \text{Loyalty_Club_Only}) + (281.84 * \text{Loyalty_Club_And_Credit_Card}) + (-245.42 * \text{Store_Mailing_List}) + (66.98 * \text{Avg_Num_Products_Purchased}) + (0 * \text{Cash_Only})$$

Step 3: Presentation/Visualization

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

I would recommend that the company send out the catalog to the 250 new customers.

2. How did you come up with your recommendation?

Apply the linear regression model to the mailing list data set to get the predicted average sale amount. Then multiply Score_Yes by Avg_Sale_Amount to get predicted revenue. After that calculate the profit = (Revenue*0.5) - 6.5

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

Profit = 21987.435