

Predicting Catalog Demand

1. Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

The task is to find out the expected revenue and subsequently the profit the company may earn from the purchases made after sending out the catalogs to the new customers.

The predictions should give insights to the stakeholders on whether to send the catalogs or not depending on if the predicted profit exceeds \$10000.

2. What data is needed to inform those decisions?

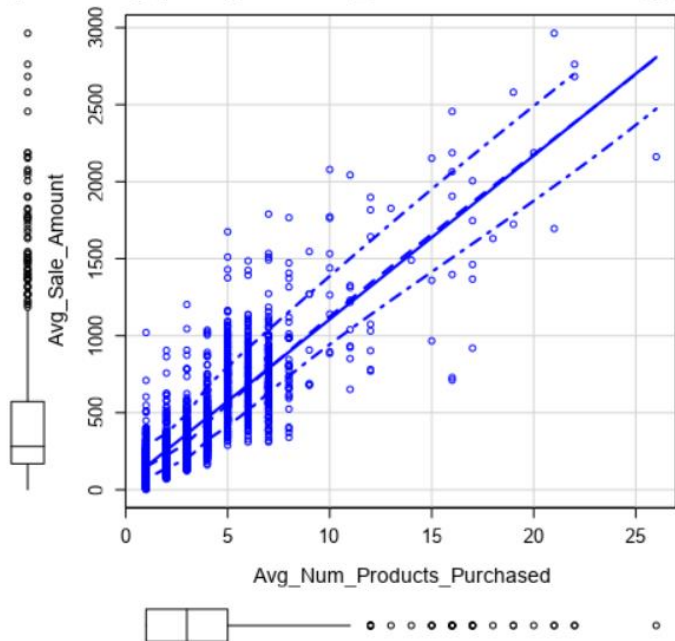
We'll need records of the previous/current customers to be used as training set and the records for the 250 customers who will be used to make the predictions on profit and aid in making a decision on whether to send the catalogs or not.

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.

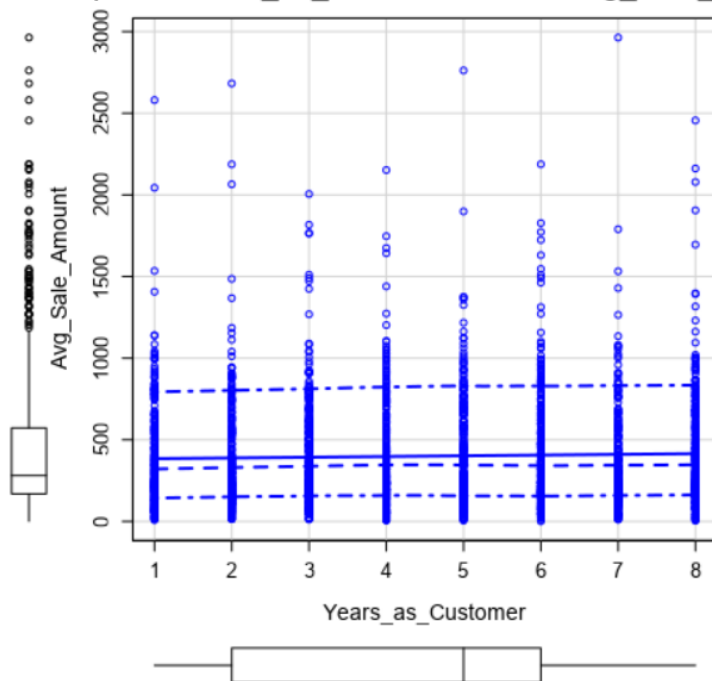
I chose Avg_num_Products_purcahsed as the continuous predictor variable to use in training the model. This is because the feature has a strong correlation with the Target variable (Avg_Sales_Amount). As can be seen from the scatterplot below, as Average number of products purchased increases, the average sales amount also increases.

plot of Avg_Num_Products_Purchased versus Avg_Sal



However, there's no discernible correlation between the other continuous variable (Years_as_Customer) and the target variable as seen from the scatterplot below.

scatterplot of Years_as_Customer versus Avg_Sale_Am



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.

My model is a good one since the Multiple R**2 value is 0.8369 meaning 83.69% of the variation in the target variable can be explained by the variation in the predictor variables. The p-value for the model is much closer to 0 meaning there's a strong relationship between my predictor variables and target variable. Below is a table showing the different metrics of variables in my model:

Report for Linear Model Predicting_Catalog_Demand

Basic Summary

Call:

lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = 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 degrees of freedom (DF), p-value < 2.2e-16

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

Projected_Sales = 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 + 0 * Credit Card Only

3.Presentation/Visualization

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

Yes, the company should proceed and send the catalog to the 250 customers

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process). In the creation of the model, features like the Years a person has been a customer and the customer segment were used to make the prediction. Using these features to get a prediction of the expected profit was deemed to be most appropriate since the correlation of these features with the average sale amount was strong. Given the fact that the company will only proceed to send the catalogs if the predicted profit exceeds \$10000 and that my prediction gives a profit of \$21987.44 which is more than

double the threshold set by the company, this is reason enough to proceed with sending the catalog.

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

\$21,987.44