

Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Provide an explanation of the key decisions that need to be made. (500 word limit)

Key Decisions:

Answer these questions

1. What decisions need to be made?

We need to determine how much profit the company can expect from sending this year's print catalog to a mailing list of 250 customers. Management does not want to mail the catalogs unless profits will exceed \$10000.

2. What data is needed to inform those decisions?

We need data on past sales to predict expected revenue from the 250 new customers. Then we need to then predict the profit by multiplying predicted sales by the probability that they will make a purchase, as well as accounting for the profit margin and the cost of printing and distributing. Then we aggregate the results to predict the total profit for the company.

Step 2: Analysis, Modeling, and Validation

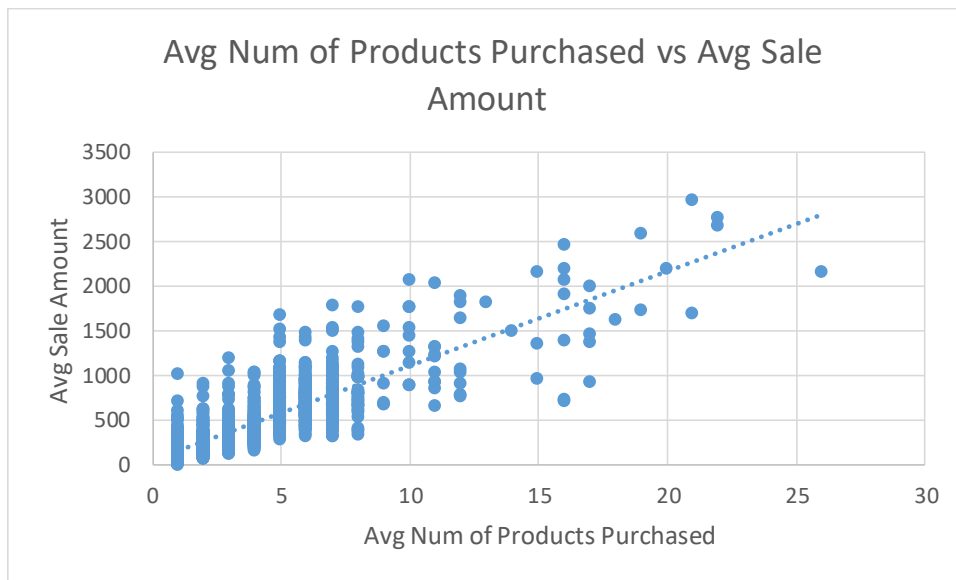
Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

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 experimented with the variables that were present in both past customer data and the new mailing list. I selected the only two variables with a p-value less than .05, meaning that a relationship exists. Num of Years as Customer just barely did not make the cut, it had a p-value just under .05 when all potential variables were included in the analysis. But when I removed all variables with high p-values, the p-value of Num of Years increased to just over .05. Number of products purchased is a numeric variable and customer segment is a categorical. The scatterplot below shows a linear relationship between customer segment and sales. We cannot graph a non-binary categorical variable (customer segment).



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.

The p-value for both predictor variables is less than .05, meaning there is likely a relationship between the predictor variable and target variable, in other words both predictor variables are statistically significant. The adjusted R-squared is value is above 0.7 so our model is considered a strong model; it should explain a lot of the variation in prices.

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)

$Y = 303.46 + 66.98 * \text{Avg_Num_Products_Purchased} + 281.84 \text{ (If Type: Loyalty Club and Credit Card)} - 149.36 \text{ (If Type: Loyalty Club Only)} - 245.42 \text{ (If Type: Store Mailing List)} + 0 \text{ (If Type: Credit Card Only)}$

Step 3: Presentation/Visualization

Use your model results to provide a recommendation. (500 word limit)

At the minimum, answer these questions:

1. What is your recommendation? Should the company send the catalog to these 250 customers? Yes, the company should send the catalog to these customers.
2. How did you come up with your recommendation? (Please explain your process so

reviewers can give you feedback on your process) I made the model using linear regression tool on Alteryx. Then I used the score tool to calculate predicted sales. Next, I multiplied these values by probability customer will respond and buy something from catalog (Score_Yes) and multiplied by 50% profit margin. Then I subtracted the \$6.50 cost of printing/distributing catalog from resulting profit value of each customer. I then summed the 250 profit values to predict the total profit.

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

\$21,987.44