

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?

The key business decision that needs to be made is whether to print the catalog and send it to the new customers and can we expect the desired profit by the sales of the items in the catalog

2. What data is needed to inform those decisions?

We would need the details of the past customer transactions and the purchase patterns of the customer with the details of the average sale amount and average products purchased

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)

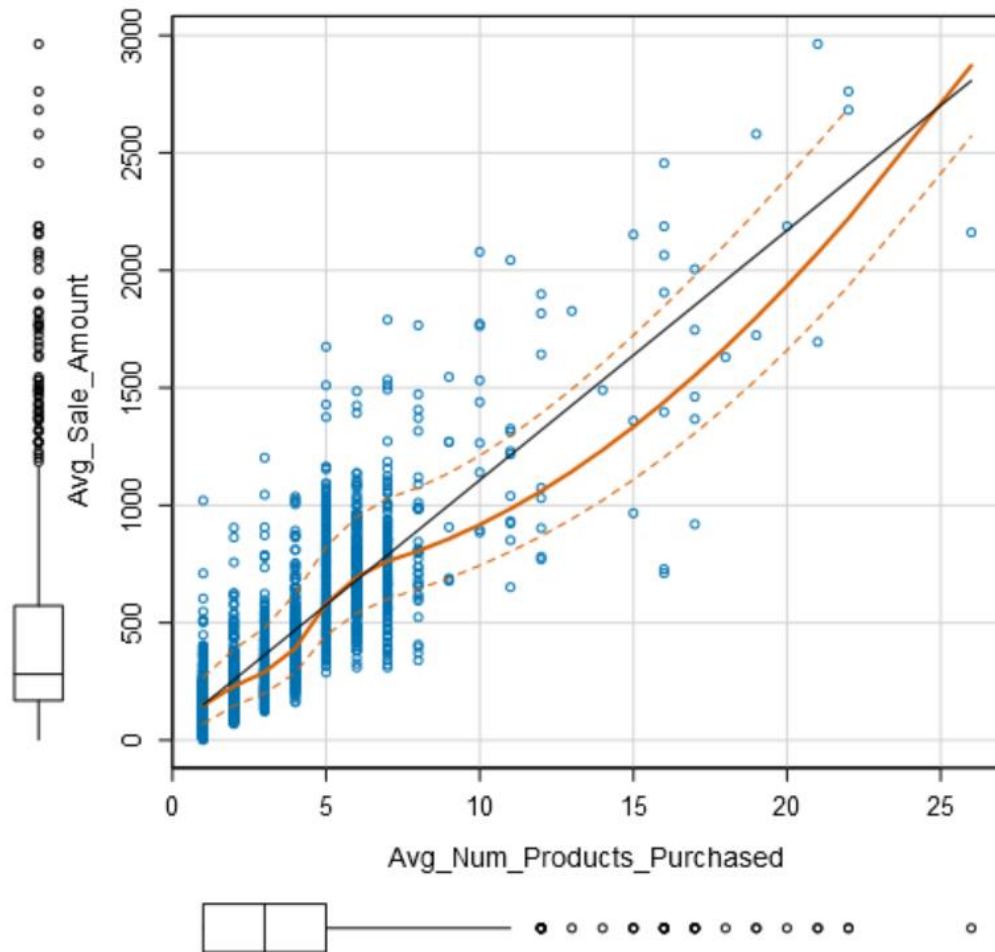
Important: Use the *p1-customers.xlsx* to train your linear model.

At the minimum, answer these questions:

1. How and why did you select the [predictor variables \(see supplementary text\)](#) in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer to this [lesson](#) to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.

Average sale amount is the target variable and the average number of products purchases is the predictor variables, Customer segment is the categorical variable. The predictor variables and categorical variables are chosen after the linear regression analysis which results in the desired R-squared and P-values

Scatterplot of Avg_Num_Products_Purchased versus Avg_Sale_Amount



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 R-squared value which I got for this model is 0.8369, which is near to 1 and hence it represents a good model and the p-value: $< 2.2e-16$ which indicated that the predictor and categorical values selected makes the new set of values a good fit into the model which has been produced

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)

Average sales amount = $303.46 + (-149.36 * \text{Customer.SegmentLoyalty Club Only}) + (281.84 * \text{Customer.SegmentLoyalty Club and Credit Card}) + (-245.42 * \text{Customer.SegmentStore Mailing List}) + (66.98 * \text{Avg.Num.Products.Purchased})$

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 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)

**The expected sales amount from the new customers is \$47224.87,
The gross profit we are expecting is 50% which is $\$47224.87 * 0.50 = \23612.43 which is more than the minimum profit the company is expecting to get from the new customers (\$10000)**

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

The company can expect a profit of $\$23612.43 - (6.5 * 250) = \21987.43 from the new catalog sales

Project Work flow in Alteryx

