

MAKE A COPY

Project 1: Predicting Catalog Demand

Complete each section. When you are ready, save your file as a PDF document and submit it here: <https://classroom.udacity.com/nanodegrees/nd008/parts/c0b53068-1239-4f01-82bf-24886872f48e/project>

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?
 - A decision needs to be made to assist in determining if sending catalogs to new customers on the mailing list will be profitable more than \$10,000.
2. What data is needed to inform those decisions?
 - Analysing the historical data from existing customers

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.
After the first review, the target variable is "avg sale amount". The predictor variables are "customer segment" and "avg sale amount". The R-squared is 0.8369 and the adjusted R-squared is 0.8366.
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 second attempts may be a good model as the R-squared is 0.7038 and the adjusted R-squared is 0.7033. Selecting the variable “customer segment” is to identify the “Store Milling List” while “Responded to last catalog” is to find who responded.

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)

Important: The regression equation should be in the form:

$$Y = \text{Intercept} + b1 * \text{Variable_1} + b2 * \text{Variable_2} + b3 * \text{Variable_3} \dots$$

For example: $Y = 482.24 + 28.83 * \text{Loan_Status} - 159 * \text{Income} + 49 (\text{If Type: Credit Card}) - 90 (\text{If Type: Mortgage}) + 0 (\text{If Type: Cash})$

Note that we **must** include the 0 coefficient for the type Cash.

Note: For students using software other than Alteryx, if you decide to use Customer Segment as one of your predictor variables, please set the base case to Credit Card Only.

The equation is:

$$Y = 303.46 + -149.36 * \text{Customer_SegmentLoyalty Club Only} + 281.84 * \text{Customer_SegmentLoyalty Club and Credit Card} + -245.42 * \text{Customer_SegmentStore Mailing List} + 66.98 * \text{Avg_Num_Products_Purchased} + 0 * \text{Customer Credit Card Only}$$

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?

Based on the results, it is recommended for the company to send the catalog to these 250 customers, as many of existing customers responded to last catalog and the expected profit exceed \$10,000.

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

I followed the giving steps by the first reviewer. First, I calculated the score of each customer by multiply [score] * [score_yes], then I calculated the summation of the previous equation. For the gross margin, I multiply the getting summation by 0.5 and then subtract the total expenses which is [6.5*250]. The final result is \$21,987

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

It is \$21,987.

Before you Submit

Please check your answers against the requirements of the project dictated by the [rubric](#) here. Reviewers will use this rubric to grade your project.