

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 needs to be made?

The decisions to be made is related to financial issue in the company to illustrate more, the result of the analysis of this project will conclude the final decision of the manager to decide regarding sending catalog to the 250 customers or not, depending on the profit that the analysis will predict.

2. What data is needed to inform those decisions?

Some historical data about the customers like their average sales, average amount number of product sales and segments and any data that impact the average sales of the companies. In addition, the cost of each catalog and the gross margin should be known to be able to calculate the profit.

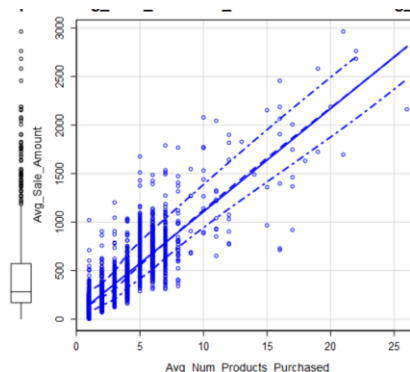
Step 2: Analysis, Modeling, and Validation

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 have selected the variable based on the impact on the average sales and the predictor was average number of products purchased and the company segment. The above graph is average sales amount vs average number products purchased. Regarding the good indicator of predictor if there is a linear relationship (i.e., whenever x increased y is increased) it means the predictor is good as the above graph. However, the categorical predictor can't be expressed as a graph so the factor that determine either it's good or not is the P-value, it should be less than 0.05 noting that if the model built on 95% confidence level.

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.

Each variable has been selected the P-value of it is less than 0.05 since the model is built on 95% confidence level the P-value is accepted and good. Furthermore, the r-squared for the model is 83.66% which is higher than 70% and this is a good indicator for the model.

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:

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

Step 3: Presentation/Visualization

1. What is your recommendation? Should the company send the catalog to these 250 customers? Since the profit exceeds 10,000 dollars which is the constraint that the manager asked for, I recommend the company to send their catalog.

3. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)
- Build a linear regression model using historical data by Alteryx.
 - Predict the sales and multiply it by the probability of score yes to get the predicted revenue.
 - Then predict profit by the following formula $(\sum \text{Expected Revenue} * 0.5) - (6.5 * 250)$.

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

The expected profit equals 21987.435 dollars