Project 1.2: Predicting Catalog Demand

By: Anthony Nguyen

Project Overview

You recently started working for a company that manufactures and sells high-end home goods. Last year the company sent out its first print catalog, and is preparing to send out this year's catalog in the coming months. The company has 250 new customers from their mailing list that they want to send the catalog to.

Your manager has been asked to determine how much profit the company can expect from sending a catalog to these customers. You, the business analyst, are assigned to help your manager run the numbers. While fairly knowledgeable about data analysis, your manager is not very familiar with predictive models.

You've been asked to predict the expected profit from these 250 new customers. Management does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000.

Project Details

- The costs of printing and distributing is $6.50 per catalog.
- The average gross margin (price - cost) on all products sold through the catalog is 50%.
- Make sure to multiply your revenue by the gross margin first before you subtract out the $6.50 cost when calculating your profit.

Step 1: Plan your Analysis

Key Decisions:
1. What decisions needs to be made?
   - - The decision to send the product catalog to the new 250 needs to be made if it is feasibly profitable by exceeding $10,000 or not.
2. What data is needed to inform those decisions?
   - - We need data on all of the customers and any data that can tell us whether they have bought something in the catalog in the past. This historical customer data needed are:
     - Average amount of products purchased
     - Average amount spent
   - - Looking at all of the scatterplots, it is clear that the Avg Sale Amount vs. Avg Number of Products Sold yielded a strong indication that the two variables are related.
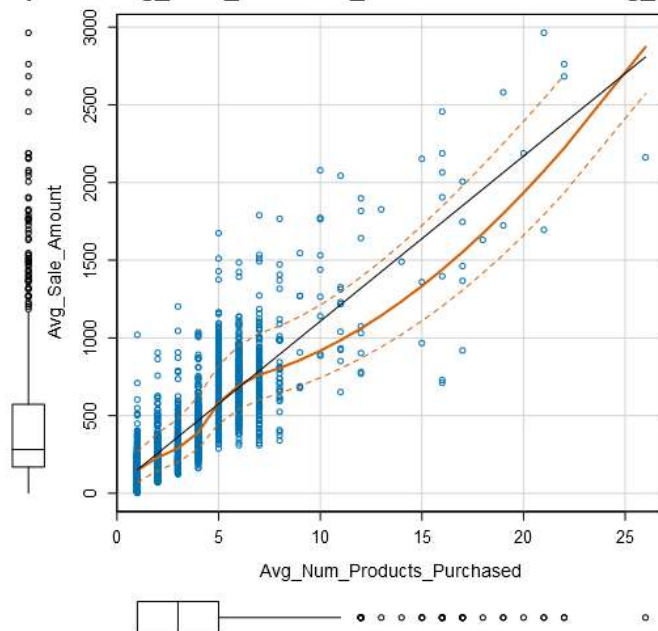
Step 2: Analysis, Modeling, and Validation

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.
   - - The way that the predictor variables were selected was initially selecting all and running the linear regression report. While reviewing the report, I discovered that many of the predictor variables were insignificant (R-Squared + P-Values were insignificant

and can be excluded from the analysis). Customer Segments and Average Number of Products Purchased were the most significant. Which makes sense because both variables help predict our target variable.

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 linear model is a good model because the predictor variables are strong as referenced with the R-Squared + P-Values. The P-Values for Customer Segment variable had *** and was far less than the 0.05 target of being insignificant. The same goes for the Avg. Number of Products Purchased segment. The Adjusted R-Squared = 0.8366. The high Adjusted R-Squared value is considered a strong model. This value represents the amount of variation in the target variable explained by the variation in the predictor variables.

	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 DF, p-value: < 2.2e-16

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)

The best linear regression equation based on the available data is as follows:
Y = 303.46 + (66.98*[Avg Num of Products Purchased]) + (-149.36*[if Cus_SegLoyaltyClubOnly]) + (281.84*[Cus_SegLoyaltyClubandCreditCard]) + (-245.42*[Cus_SegStoreMailingList]) + (0 * [Cus_SegCreditCardOnly])

## Step 3: Presentation/Visualization

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

My recommendation is to plan to implement the catalog distribution to the customers is a strong option for potential profit and growth for the company. 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 justification of my recommendation is as follows:

1. After analyzing the linear regression report, the R-Squared and P-Values of the target variable (Avg Sale Amount) and predictor variables (Customer Segment + Avg Number of Products Sold) confirmed to be significant.
2. Determining the final expected profit to be $21,987.44 = ([Score] * [Score_Yes]*0.5)-6.5
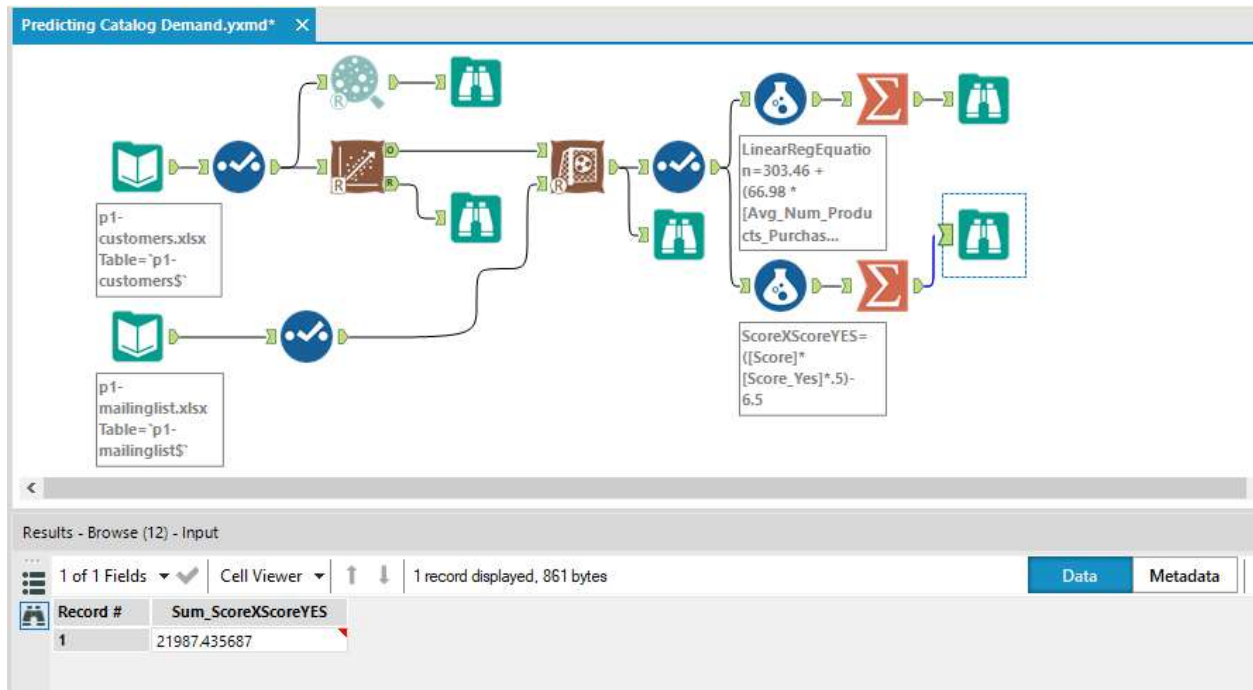
3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?

The expected profit from the new catalog distribution to the 250 customers is $21,987.44. This is taking into account the average gross margin as well as the cost of printing and distribution.