Predicting Catalog Demand

Christian Guzman

## **Step 1: Business and Data Understanding**

A company manufactures and sells high-end home goods. Last year it sent out its first print catalogue and is preparing to send out the new catalog for the year. From the mailing list the company has 250 new customers they would like to send the catalogue to. The company wants to determine how much profit can be expected from sending the catalogue to these customers. Management has set a profit threshold of $10,000. If the expected profits meet or exceed this threshold, then the catalogue will be sent.

### **Key Decisions:**

1. What decisions needs to be made?

Whether or not the print catalog should be sent to the 250 new customers within the mailing list. This is dependent on whether the estimated profit meets and/or exceeds $10,000. If the value is below $10,000, the catalog will not be sent to those 250 customers.

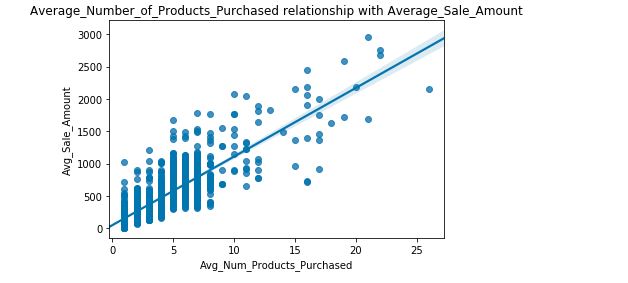
2. What data is needed to inform those decisions?

* For each of the 250 new customers we will need to estimate the average sale amount.
* Then the expected revenue can be calculated by multiplying average sale amount with the score value metric, which acts as the probability the customer will respond to the catalog.
* The average gross margin is 0.5. The price of printing and sending each catalog is $6.50 so this is then subtracted from the expected revenue times the average gross margin.
* These values are then summed to give the final expected profits.

## **Step 2: Analysis, Modeling, and Validation**

**1. How and why did you select the predictor variables in your model?**

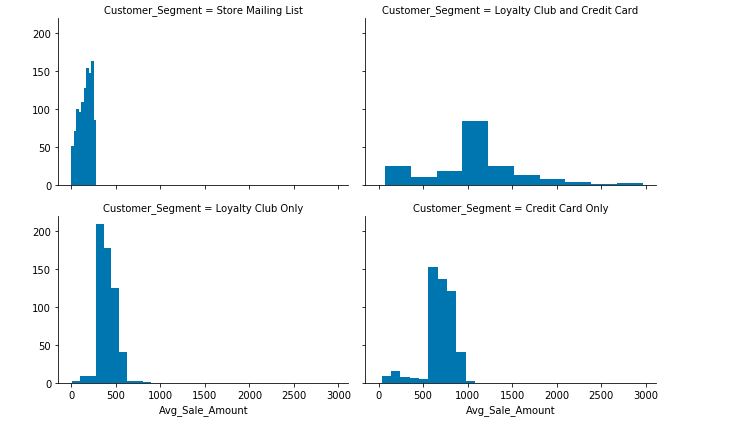
The predictor variables chosen were **Avg\_Num\_Products\_Purchased** and **Customer\_Segment**.

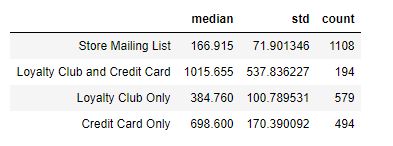


The above scatterplot clearly shows that as the average number of products, the higher the average sale amount tends to be. This makes intuitive sense as the more products bought increases the sale total.



Customers segmented under Loyalty Club and Credit Card tend to have the highest average sale amounts, while those in Store Mailing List tend to have the lowest among the 4 groups. This suggests that the customer segments may have some correlation with average sale amount and can be used as a significant predictor variable.

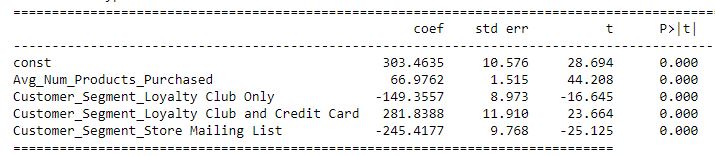




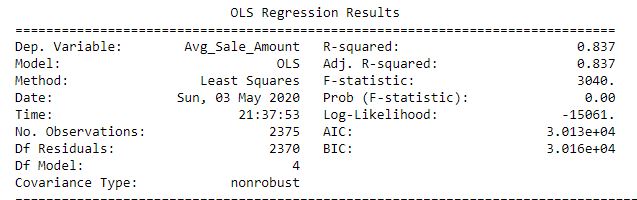
It can be gathered from the above graphs and table that, customers in the Loyalty Club and Credit Card segment, tend to spend the most, the median value was $1015 this group also had the highest variance in spending with a standard deviation of $537. The other groups spent less on average and the differences in spending were much less, and had less variance.

With this categorical variable the model should be able to take into account the customer segment when predicting for each new customer in the mailing list. It can be expected that the linear model will predict that customers in the Store Mailing List will usually spend less than those in other segments.

**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.**



From the results, All of the chosen variables are statistically significant. (P-values reported as zero)



The Adjusted R-squared is 0.837, the rule of thumb is that a value higher than 0.7 signifies a good model. In this case over 80% of the variance is explained by the model.

**3. What is the best linear regression equation based on the available data?**

Y = 303.46 + 66.98 \* Avg\_Num\_Products\_Purchased

– 149.35 \* Customer\_Segment\_Loyalty\_Club\_Only

+ 281.84 \* Customer\_Segment\_Loyalty\_Club\_and\_Credit\_Card

– 245.42 \* Customer\_Segment\_Store\_Mailing\_List

## **Step 3: Presentation/Visualization**

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

My recommendation is that yes, the company should send the catalog.

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

The estimated profit is $21987.44 This exceeds the threshold of $10,000 set my management by nearly $12,000.

Steps taken to determine the estimated profit:

1. I used predictor variables: average number of products purchased and customer segment to fit a model predicting for average sale amounts for each new customer in the mailing list.
2. The estimated sale amounts were then multiplied by the average gross margin, 0.5 and the probability the customer would make the purchase, to get expected revenues.
3. The cost of printing and sending each new catalog is $6.50, so this is subtracted from the expected revenue to get the expected profits for each customer.
4. The expected profits are then aggregated to get the total: $21987.44.

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

The model estimates that if the new catalog is sent, the total expected profit will be approximately $21K.