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? Does the company need to send the new catalog to the 250 customer? How much profit the company would make by sending catalog to the new customers.

2. What data is needed to inform those decisions?
    1. Sales from existing customers
    2. Cost of printing and distributing the catalog
    3. Av gross margin

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:

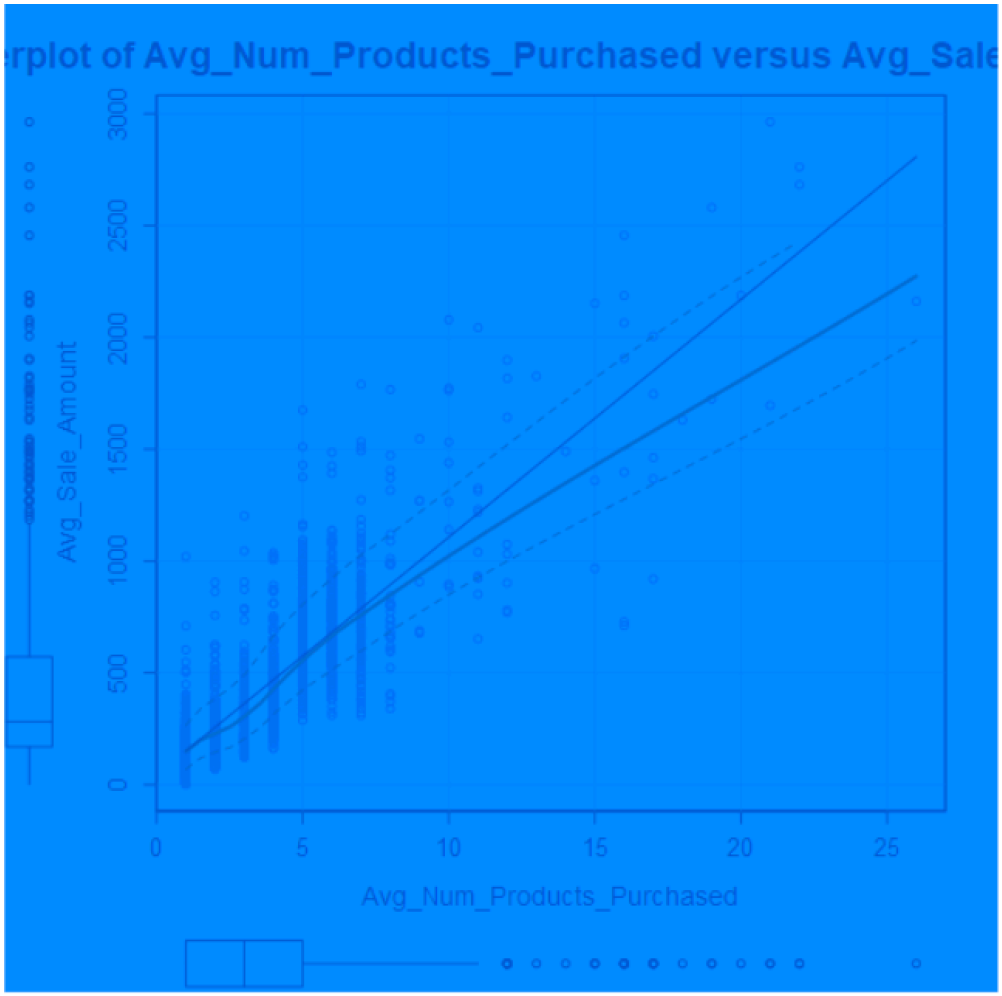
1. How and why did you select the predictor variables (see supplementary text) in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer to this lesson to help you explore your data and use scatterplots to search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

I was looking at 3 variables:

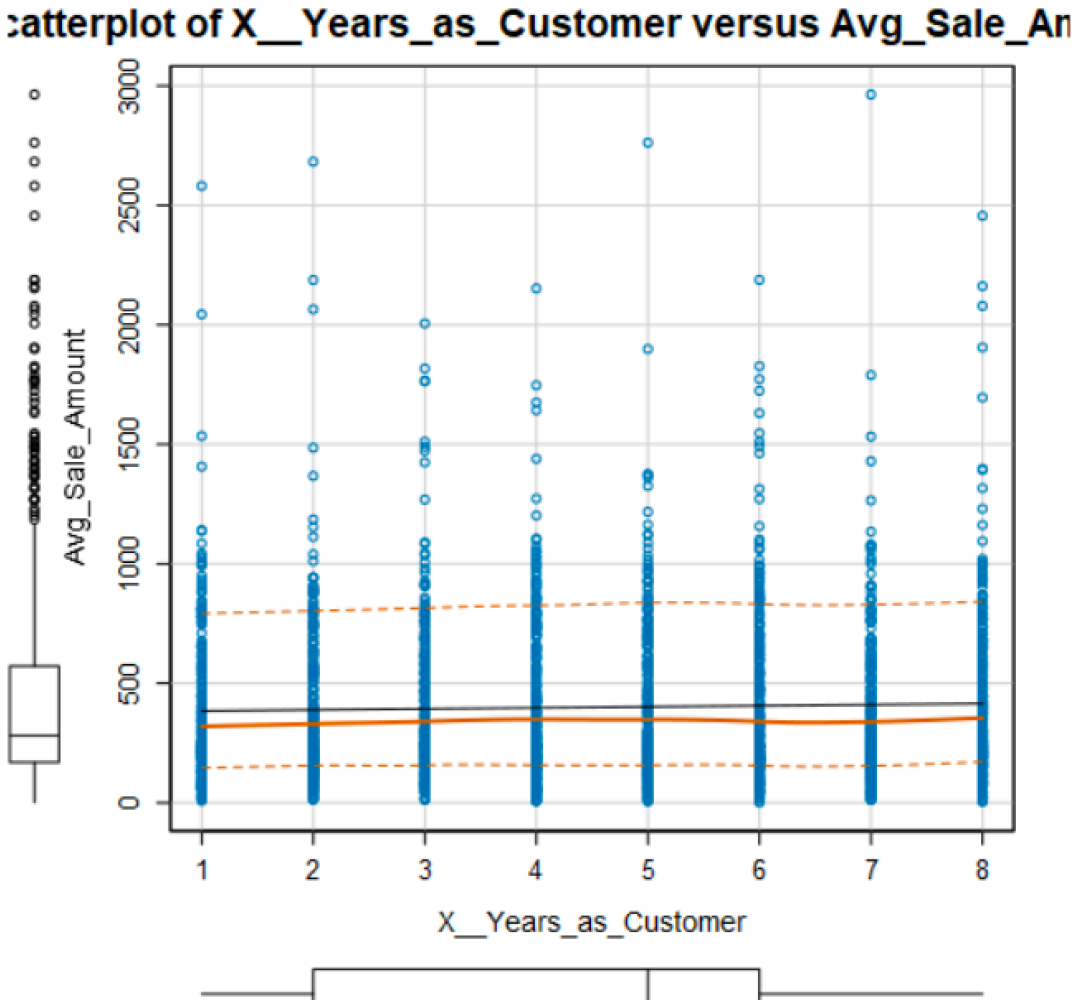
1. avg_num_products_purchased
2. # of years as customer
3. customer segment.

In relationship to the target variable – avg_sales_amount.

I noticed that sales increased linearly with product sold, which is kind of expected.

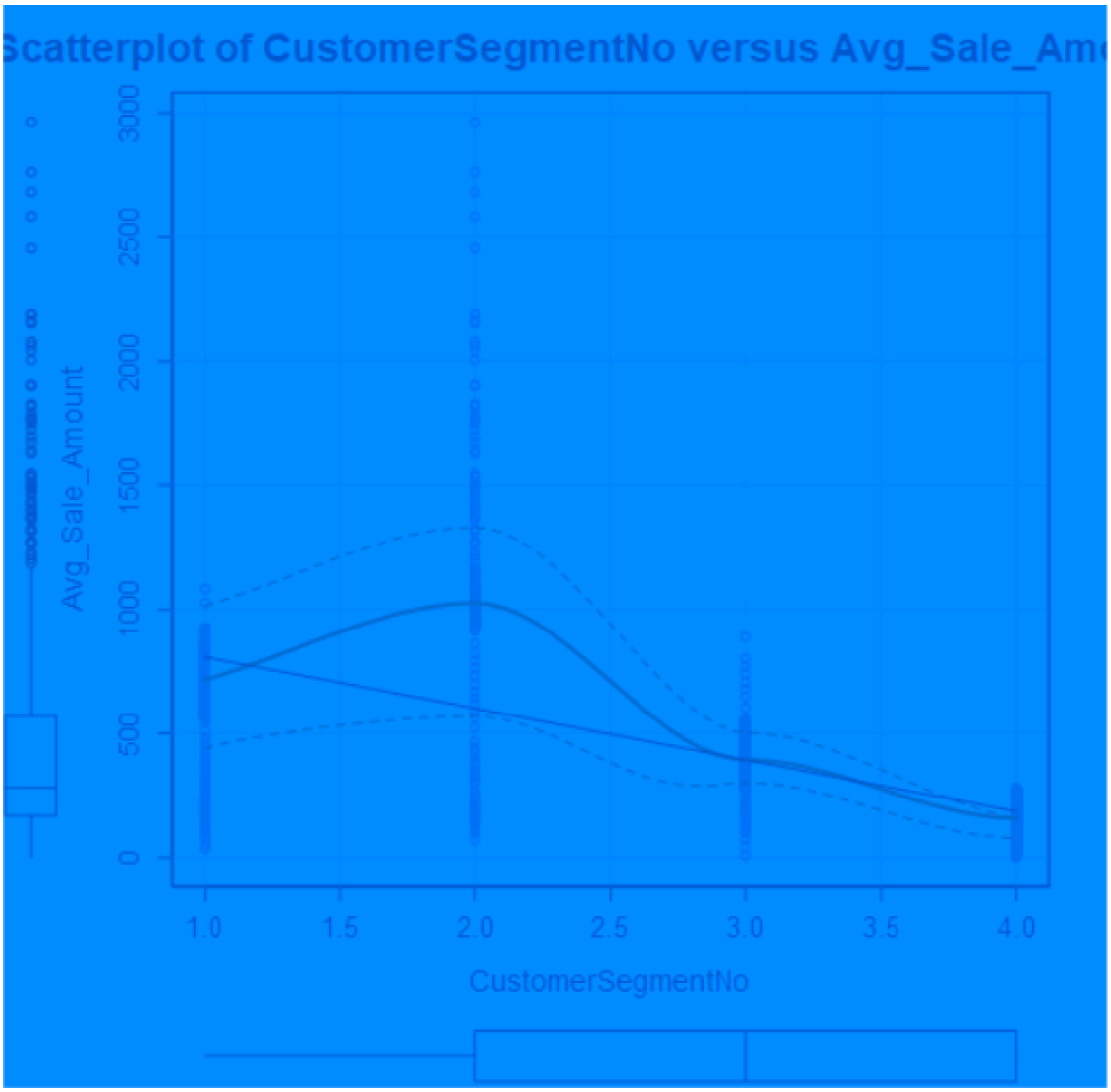


However, # of years as customer did not impact the sales. The sales kind of had similar patter regardless of years of being a customer. This means new customers contribute to the sales as much has returning customers.



There are 4 customer segments in the data and I gave them a numerical value to see how segments could have impacted sales. Following were the numbers assigned:
1 – Credit Card Only
2 – Loyalty Club and Credit Card
3 – Loyalty Club Only
4 – Store Mailing List

Sales were high for Loyalty Club and Credit Card. It was similar for “Credit Card Only” and “ Loyalty Club Only”. It was low for Store Mailing List.



Since the number of years being a customer did not impact, I dropped that variable and chose 2 predictor variables: avg_num_products_purchased and customer segment to train the model.

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.

I tried to model with the 2 predictor variables - : avg_num_products_purchased and customer segment and obtained the following:

a. p-value < 2.2 e-16
b. Multiple R-squared : 0.8369 and Adjusted R-squared :0.8366

I then removed avg_num_products_purchased and obtained the following:

c. p-value < 2.2 e-16
d. Multiple R-squared : 0.7024 and Adjusted R-squared: 0.702

Since the adjusted R square improved from 0.702 to 0.8366, avg_num_products_purchased helped improve the model.

Since p value is pretty low, the relationship between the predictor variables and the target variable is statistically significant.

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)

Av Sales = 303.46 - 149.36(If CustomerSegment : Loyalty Club Only) + 281.84 (If CustomerSegment : Loyalty Club and Credit Card) -245.42 (If CustomerSegment : Store Mailing List) + 0 (If CustomerSegment : Credit Card Only) + 66.98 * 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?

The company does not need to send catalog to the new customers.

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

The predicted revenue from the 250 new customers is approx $47,225. If we take into consideration the printing and distribution of the 250 catalog and the gross margin, we get the net predicted profit to be approx $21,987. The company mentioned it is ready to send catalog if the profit is > $10,000. Though the profit will be greater that what the company is anticipating, it still need to send out the new catalog to attract the customers. From the existing data we see customers belonging to Loyalty and Credit card segment made more purchases. The mailing list segment had the least number of customers. Also, from the exiting data, only 171 users responded to the last catalog while 2204 users purchased without responding. It seems like catalog is not impacting the customer's decision to purchase the product.

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

If the company does send across the catalog, the expected profit is $21,987

Extra Credit

1. Show the distributions for each variable in the Customer List dataset. How would these distributions affect your analysis? Would you need to go back to your manager and try to get more data?

   In the above scatter plots, we can see how number of years of being a customer did not impact the sales. Also, the customer segment was higher for credit card and loyalty club made more purchases. Customers who are part of the store mailing list and who might be getting more information about the products do not seem to increase the sales. Almost 93% purchases happened from customers who did not respond to the last catalog. I would like to understand the marketing strategies of the product. Do we have

   a. Referral by friends
   b. Promotion code on sites
   c. Cross recommendation of the products on the sites

which are impacting the sales. These information would help understand what do user respond to and the conversion rate of making a user purchase the product. This can help formulate the strategies to increase the user base and ultimately revenue.

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.