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 decision that need to be made is whether to send the catalog to the 250 customers in the dataset based on the calculated expected profit.

2. What data is needed to inform those decisions?

Management will only send the catalogs if the expected profit exceeds $10,000. In order to make this decision we would need to figure out the expected revenue from each of the customers and then account for all costs associated with printing and mailing a catalogue. We will also need any customer data that can tell us whether they've bought something in the catalogue in the past, including but not limited to: 1) bought an item from a past catalogue 2) average amount of items the customer buys from the company 3) the total dollar amount that the customer spent ordering from the catalogues.

Step 2: Analysis, Modeling, and Validation

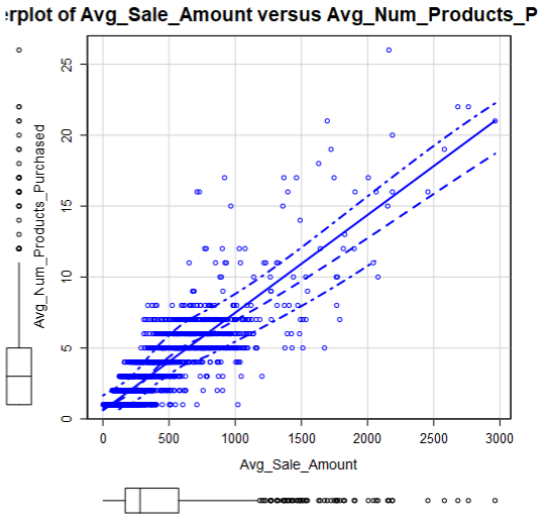
Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

Important: Use the p1-customers.xlsx to train your linear model.

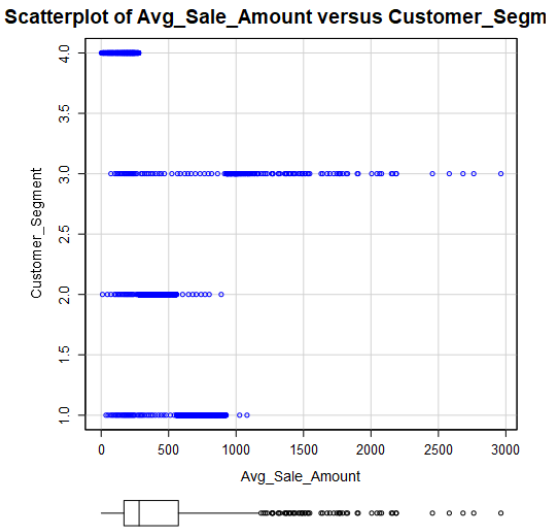
At the minimum, answer these questions:

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.

    The target variable for this linear regression model is [Avg_Sale_Amount]. The predictor variables used for this linear regression model is [Customer_Segment] and [Avg_Num_Products_Purchased]. I chose the predictor variables based on their p-value being less that 0.05. The p-value for both variables is < 2.2e-16 which tells us that these variables are statistically significant.



The scatter plot show above between avg_sale_amount and avg_num_products_purchased shows that as the avg_num_products_purchased increased, so does the avg_sale_amount. We can see there is a positive linear relationship between the two variables. This is consistent with the regression model output since the p-value is <0.05.



The scatter plot show above between avg_sale_amount and customer_segment shows that as certain types of customers can be associated with a higher avg_sale_amount. We can see there is a relationship between the customer segments and the avg_sale_amount. Customer_segmentLoyalty Club and Credit Card tends to result in an increase in avg_sale_amount. This is consistent with the regression model output since the p-value is <0.05.

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 predictor variable selected for the linear regression model is statistically significant based on the p-values being lower than 0.05. The linear regression model also produces a high adjusted R-squared value of 0.8366 which indicates a strong model.

Coefficients:

	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 degrees of freedom (DF), p-value < 2.2e-16

Type II ANOVA Analysis

Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)	
Customer_Segment	28715078.96	3	506.4	< 2.2e-16	***
Avg_Num_Products_Purchased	36939582.5	1	1954.31	< 2.2e-16	***
Residuals	44796869.07	2370			

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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)

Y= 303.46 - 149.36*[Customer_SegmentLoyalty Club Only] +
281.84*[Customer_SegmentLoyalty Club and Credit Card] –
245.42*[Customer_SegmentStore Mailing List] + + 0*[Customer_SegmentCredit 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?

It is recommended that the company sends the catalog to these 250 customers, because the expected profit exceeds $10,000.

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

The recommendation was concluded by calculating the expected profit for sending the catalog to these 250 customers. In order to calculate the expected revenue, I used the Score tool which predicted the avg_sale_amount for the 250 customers. In order to get expected revenue I multiplied the Score by the probability that the customer will respond to the catalog (Score_Yes).

The expected profit was calculated using the below formulas.

- ➔ expected_rev = [Score_Yes] * [Score]
- ➔ expected_profit = ([expected_rev]* 0.5) – 6.5

In the expected profit formulate the 0.5 refers to the average gross margin of 50% and the $6.5 refers to the cost of printing and distributing the catalogue.

Then I calculated the total sum of all the expected_profit for all 250 customers to find the total expected profit from sending the catalog to these 250 customers.

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 for these 250 customers is $21,987.43.