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?

decide whether the catalog should be sent to new customers or not based on estimated profit.

2. What data is needed to inform those decisions?
   1- calculate the average number of sale amount per customer(Avg_Num_Products_Purchased).
   2- Information about current customers (customers dataset).
   3- Information about new customers (mailinglist dataset).
   4- Profit margin = 50%
   5- Catalogue cost = 6.50

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?

the average number of products purchased and customer segment are the predictor variables. only these variables have the p-value less than 0.05 which shows statistically significant. Here is a scatterplots between Avg_sales_amount vs customer_segments and avg_num_of_products_purchased:

tterplot of Avg_Num_Products_Purchased versus Avg_Sale_

Avg_Sale_Amount

Avg_Num_Products_Purchased
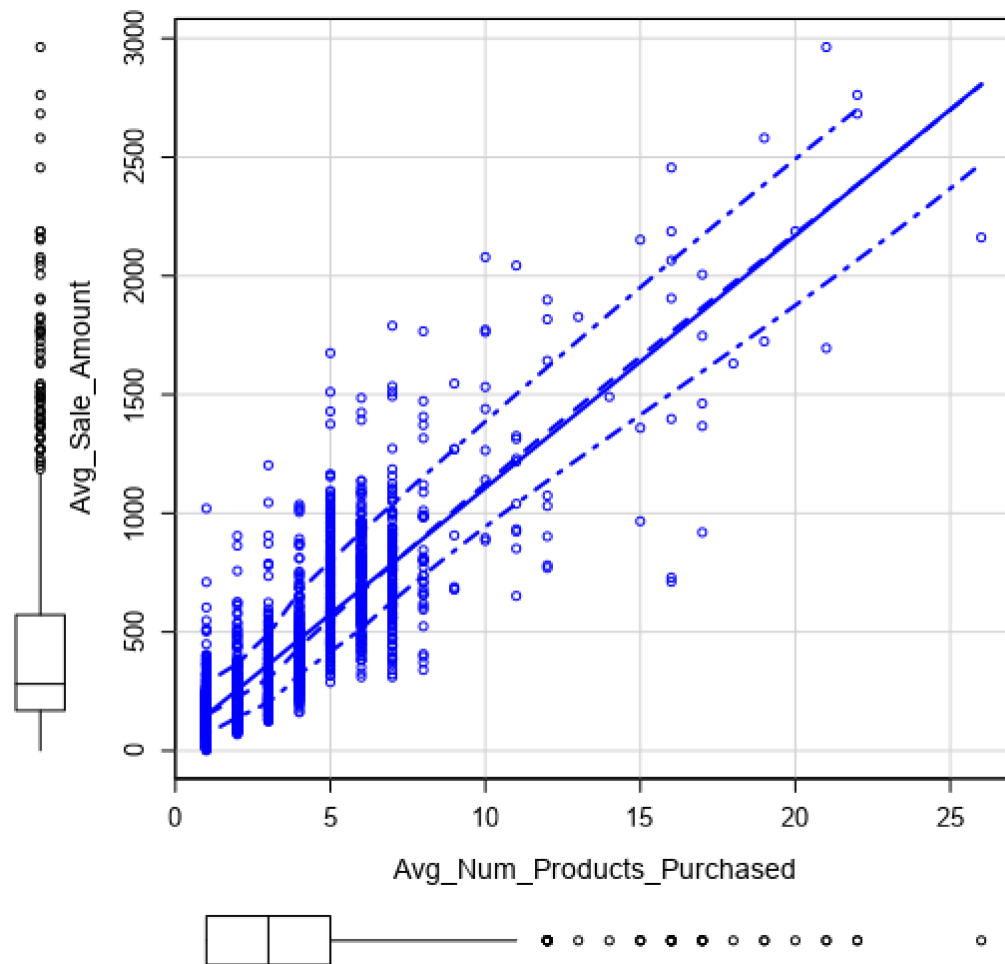
Scatterplot of Avg_Sale_Amount versus Customer_Segmer
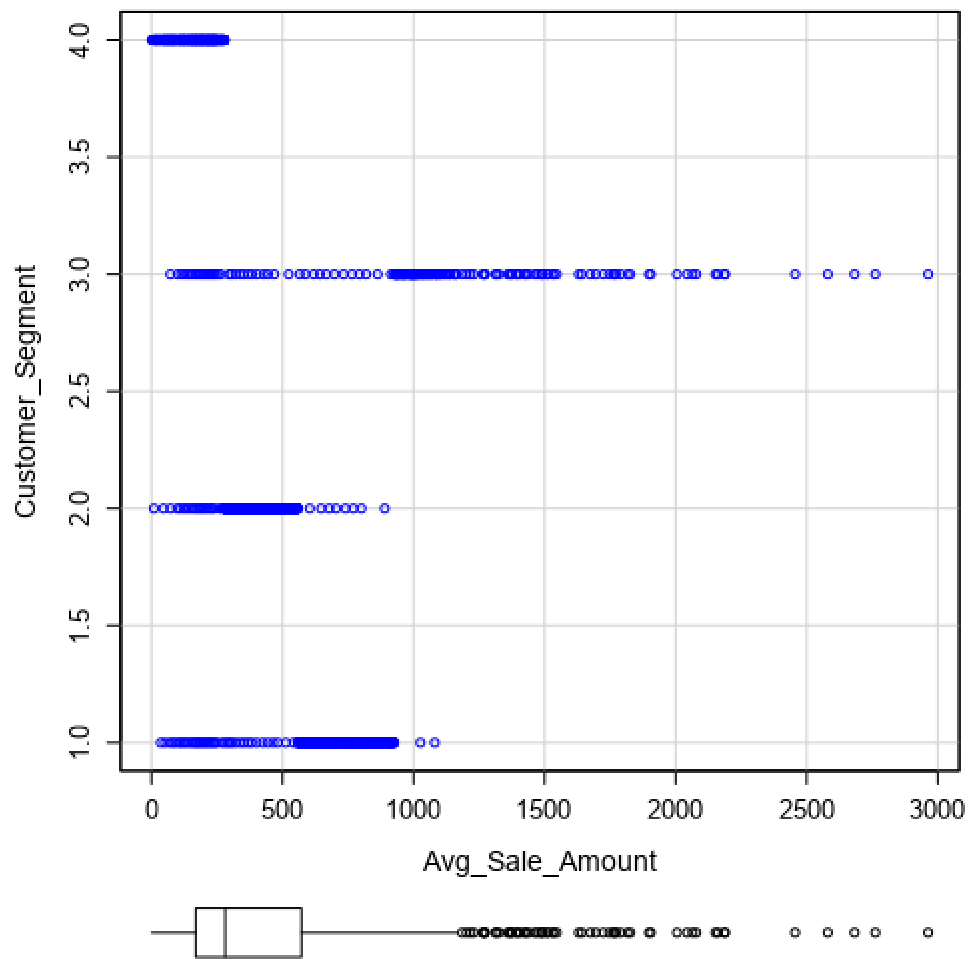
customer segment is categorical variable and is separated by 4 categories.

1. Explain why you believe your linear model is a good model.

Record	Report
1	**Report for Linear Model linear_regression**
2	*Basic Summary*
3	Call:
	lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)
4	Residuals:

Min	1Q	Median	3Q	Max
-663.8	-67.3	-1.9	70.7	971.7

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

Our predictor variables have p-values less than 0.05 and the Adjusted R Squared value is 0.8366 which is statistically significant. Its mean this is good model.

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

Avg_Sale_Amount = 303.46 - 149.36 * (Loyalty Club Only) + 281.84 * (Loyalty Club and Credit Card) - 245.52 * (Store Mailing List) + 66.984 * (Avg_Num_Products_Purchased) + 0 (if credit only).

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?

Since the condition was that if the profit exceeds $10000, and it actually exceeds as calculated using linear regression model, so yes the company should send these catalogues to these 250 customers.

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

1- calculated Avg_Sales using the linear regression model
2- Created a new column (predict_avg_Sales = Avg_Sales * Score_Yes)
3-profit_margin = 50%
4- cost of each catalogue = 6.50
5- profit = predict_avg_Sales *$0.5-(6.50*250)$

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

Profit = $21987.435687

Alteryx Workflow :