Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

Answer these questions

1. **What decisions needs to be made?**

   The decision to be made for sending catalog to 250 customers. If the predicted profit from this action greater than $ 10,000 then we should send catalog.

2. **What data is needed to inform those decisions?**

   Sales data, number of products purchased, customers segments.

Step 2: Analysis, Modeling, and Validation

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

I ran linear regression study across most/all variables against *Avg_Sale_Amount to* see which variables are likely related to the Average Sale Amount and have linear relationship. As shown in figure no.1 we can see Customer_Segment & Avg_Num_Products_Purchased have statistical significance which mean that the results did not occur randomly.

Moreover, scatterplots are made for better understanding of the chosen predictor variables.

Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)	
Customer_Segment	28750472.02	3	507.42	< 2.2e-16	***
Customer_ID	4116.78	1	0.22	0.64063	
ZIP	13488.52	1	0.71	0.39814	
Store_Number	19045.4	1	1.01	0.31539	
Avg_Num_Products_Purchased	36828004.72	1	1949.95	< 2.2e-16	***
X._Years_as_Customer	69384.42	1	3.67	0.0554	.
Residuals	44685898.37	2366			

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

Figure 1

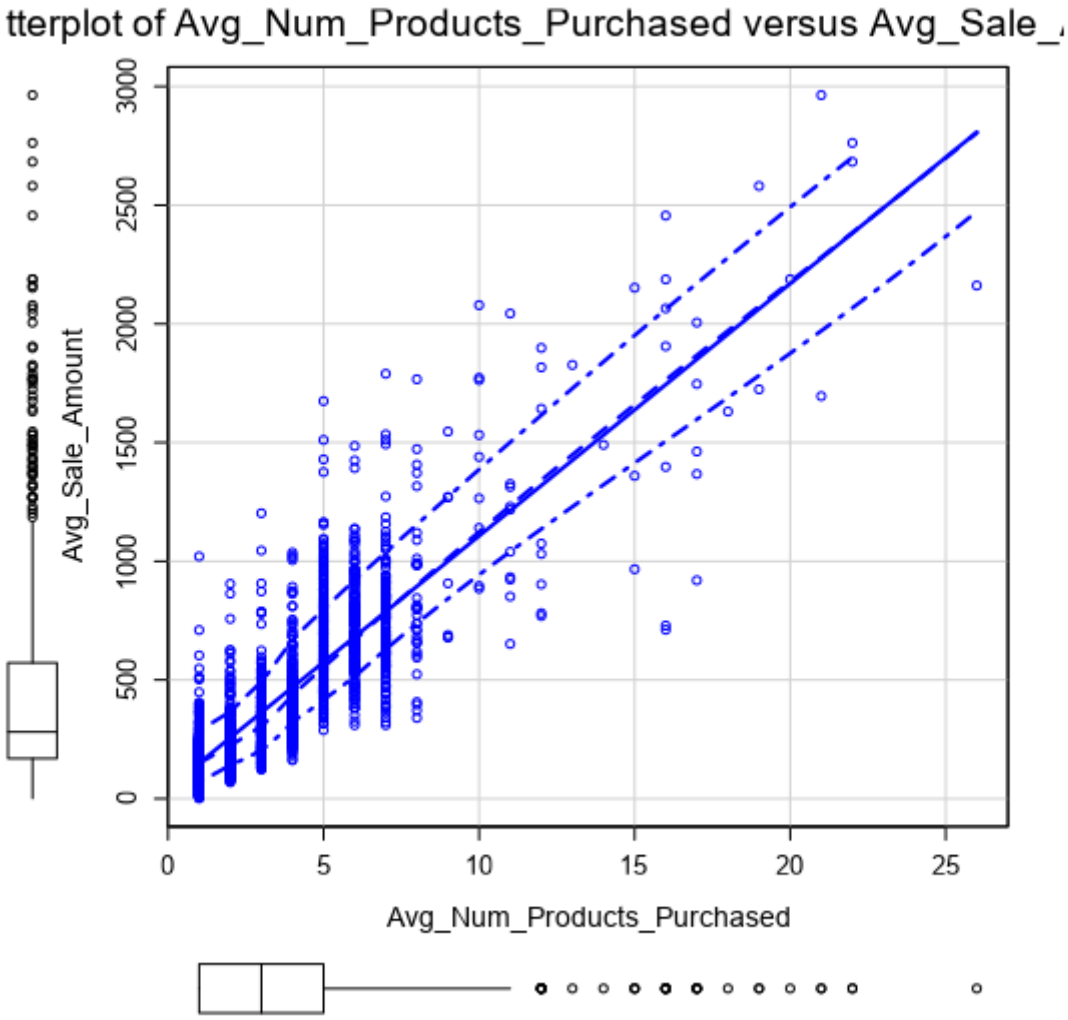
tterplot of Avg_Num_Products_Purchased versus Avg_Sale_
Avg_Sale_Amount
0
500
1000
1500
2000
2500
3000
0
5
10
15
20
25
Avg_Num_Products_Purchased

Figure 2

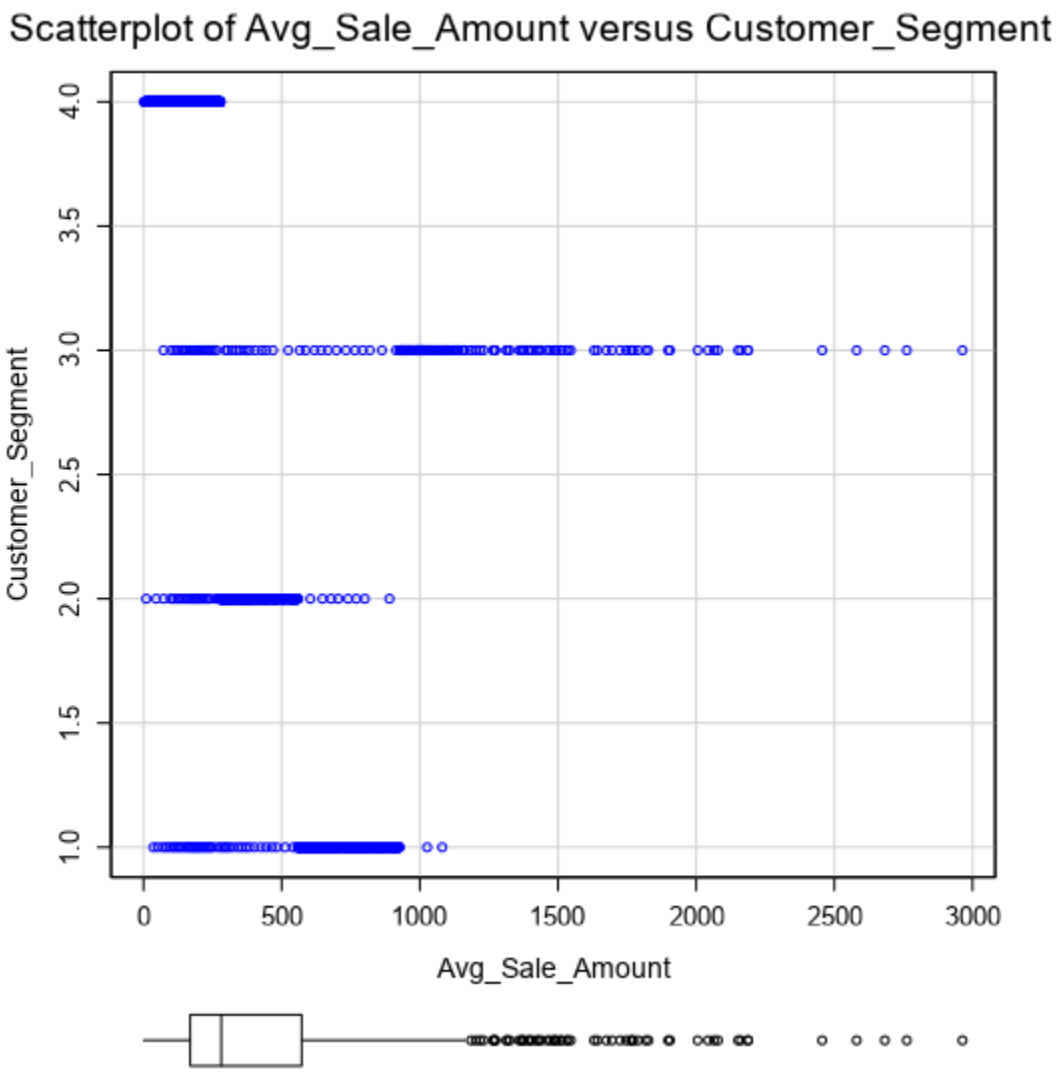

Figure 3

2. **Explain why you believe your linear model is a good model.**

   Both Customer_Segment & Avg_Num_Products_Purchased have a p-value lower than 0.05, which mean that they are statistical significance and have strong relationship with the predictor and target variable.

   In addition, we have R-squared of 0.8366 and its considered high value which is good for our model.

Record	Report
1	**Report for Linear Model Linear_Regression_7**
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

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

Avg_Sale_Amount = 303.46 – 149.36 * (If Type: Loyalty Club Only) + 281.84 * (If Type: Loyalty Club and Credit Card) – 245.42 * (If Type: Store Mailing List) + 0 * (If Type: Credit Card Only) + 66.98 * (Avg_Num_Products_Purchased)

Step 3: Presentation/Visualization

Use your model results to provide a recommendation. (500 word limit)

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

   Yes, I would recommend sending the catalog 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)

   - By using the linear regression model and applying the equation, I predicted the sales for every customer from the 250 list.
   - The predicted sales multiplied by Score_Yes values.
   - All these values of sales are sum and that's make the revenue.

     Then, the profit is manually calculated. The profit is greater than our condition (Profit > $ 10,000).

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

   Sum of expected revenue = $ 47,225.87
   Gross Margin = 50% ~ 0.5
   Cost of catalog = $ 6.50 per catalog

   = (47,225.87 * 0.5) – (6.50 * 250)

   = 23,612.44 – 1625

   = ***$ 21,987.44*** is the expected profit

Workflow:-

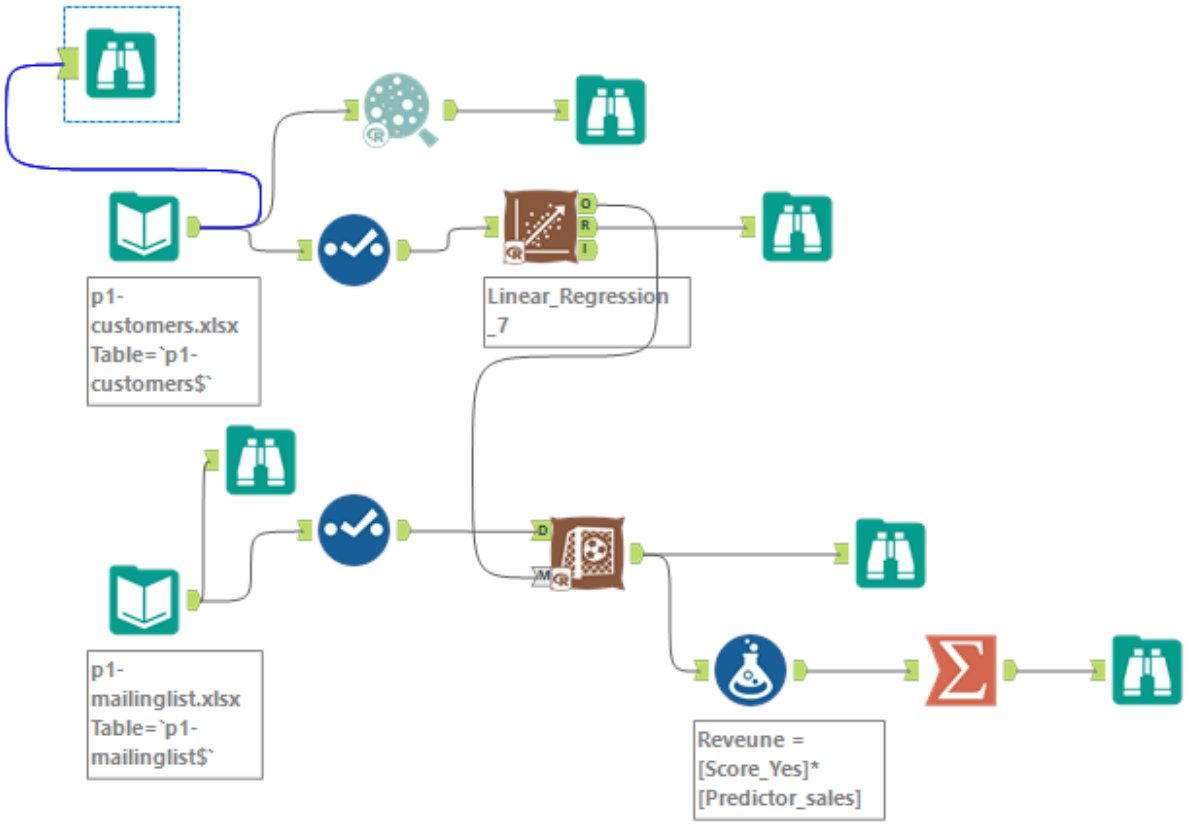

Figure 4: Workflow