Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

Should the company send the catalog to these 250 new customers based on expected profit?

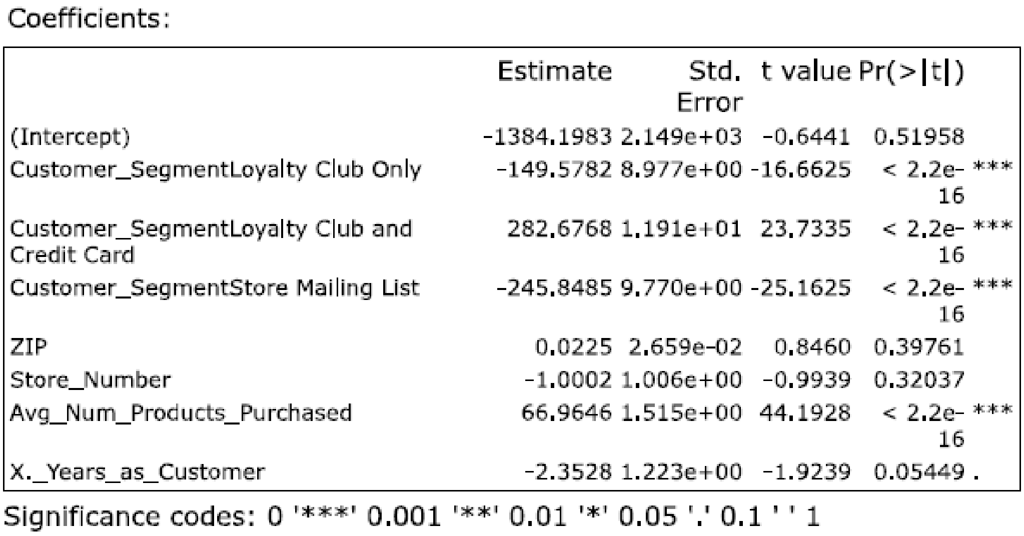
2. What data is needed to inform those decisions?

The data needed to predict sales and calculate expected profit are *Customer Segment, Average Number of Product Purchased, _ScoreYes, the average margin* and *Cost of Catalog*.

Step 2: Analysis, Modeling, and Validation

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

First, I performed a linear regression with all variables against Average Sale Amount.



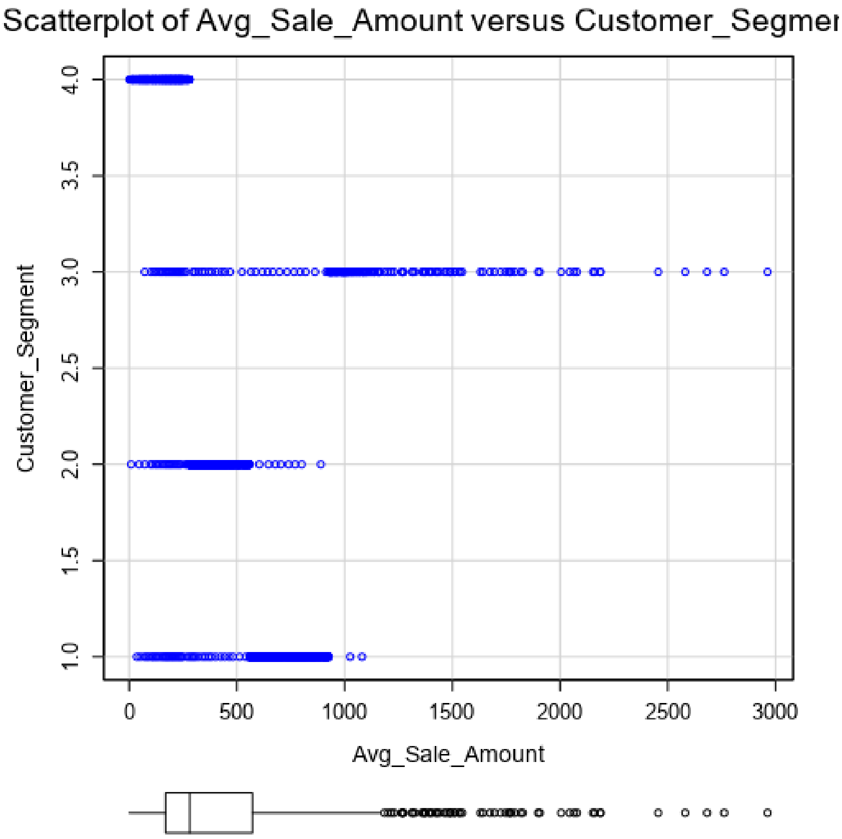
Coefficients:

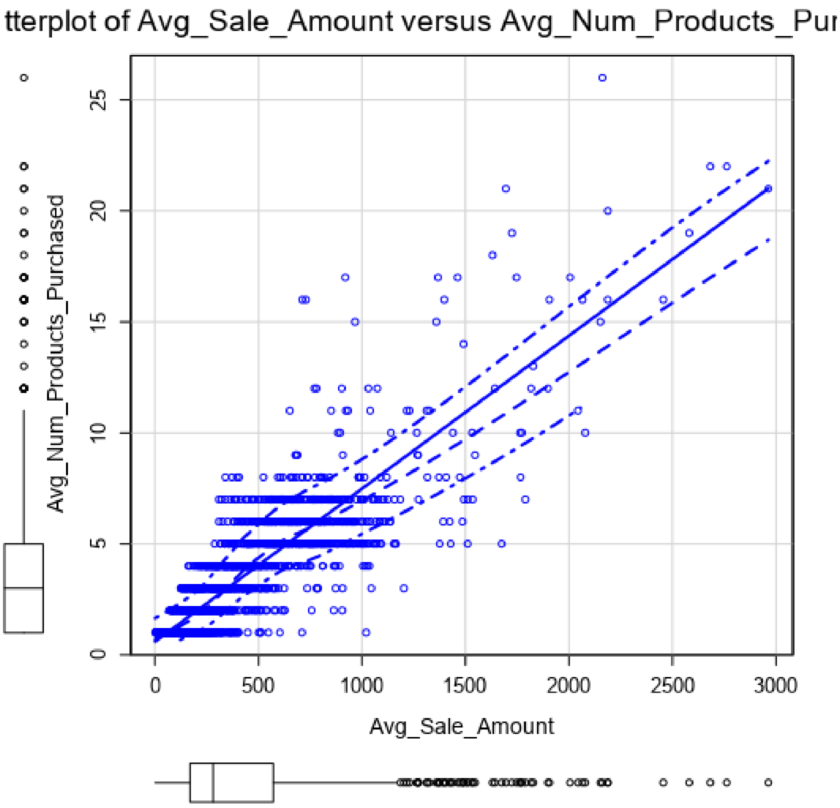
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1384.1983	2.149e+03	-0.6441	0.51958	
Customer_SegmentLoyalty Club Only	-149.5782	8.977e+00	-16.6625	< 2.2e-16	***
Customer_SegmentLoyalty Club and Credit Card	282.6768	1.191e+01	23.7335	< 2.2e-16	***
Customer_SegmentStore Mailing List	-245.8485	9.770e+00	-25.1625	< 2.2e-16	***
ZIP	0.0225	2.659e-02	0.8460	0.39761	
Store_Number	-1.0002	1.006e+00	-0.9939	0.32037	
Avg_Num_Products_Purchased	66.9646	1.515e+00	44.1928	< 2.2e-16	***
X._Years_as_Customer	-2.3528	1.223e+00	-1.9239	0.05449	.

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

As shown above, only Average Number of Product and Customer Segment have a p-value of less 0.05 which implies statistical significance.

Then, I plotted the scatterplot for each of them to see the linearity.

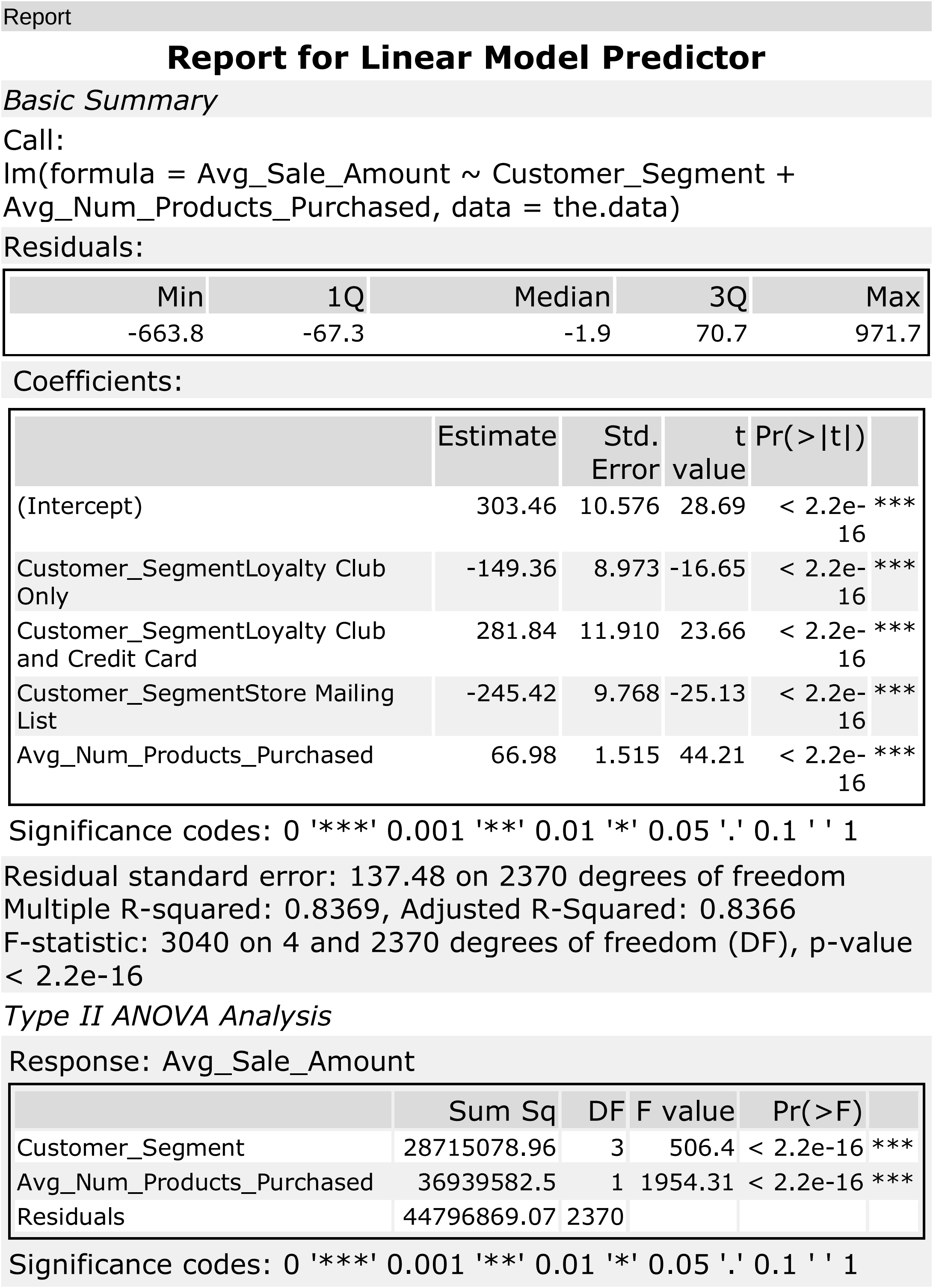




Therefore, *Customer Segment* and *Average Number of Product Purchased* will be the variables used in the model.

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

With the Alteryx linear regression function, the adjusted R-squared value of 0.8366 were calculated, which is a high value.
Customer Segment and Average Number of Products also have a p-value lower than 0.05, implying their statistical significance. Thus, the model should be considered a good model.



Report

Report for Linear Model Predictor

Basic Summary

Call:
lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)

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 x (If Type: Loyalty Club Only) + 281.84 x (If Type: Loyalty Club and Credit Card) – 245.42 x (If Type: Store Mailing List) + 0 x (If Type: Credit Card Only) + 0 x (If Type: Cash) + 66.98 x (Avg_Num_Products_Purchased)

Step 3: Presentation/Visualization

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

Yes, the company should send the catalog to these 250 new customers.

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

With linear regression model above, the possible revenue from each customer is determined by multiplying expected sale amount with Score_Yes value, then multiply by 0.5 since the gross margin are 50%. And the final profit will be the possible revenue we have subtracting $6.5 per catalog.

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

Expected profit will be 22057.