Predicting Catalog Demand

Step 1: Business and Data Understanding

What decisions need to be made?

The company wants to send catalogs to new customers but it has to be predicted whether those cutomers will be interested in catalog or not. If and only if, the profit is predicted to be more than $10,000, then only catalogs are to be send.

What data is needed to inform those decisions?

We are assigned with two datasets. In the first one, we have a list of 2,300 customers and average sales generated from them.
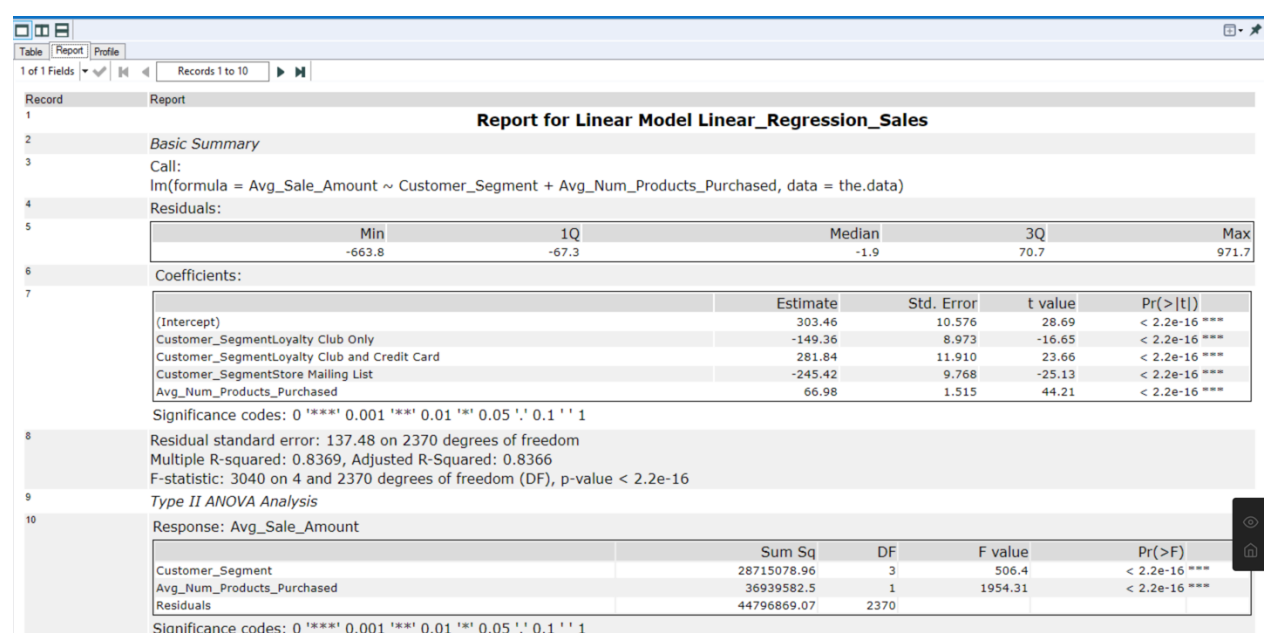 **p1-customers.xlsx** - This dataset includes the following information on about 2,300 customers
p1-mailinglist.xlsx - This dataset is the 250 customers that you need to predict sales. This is the list of customers that the company would send a catalog to.

Only when the profit will be above $10,000, company will send the catalogs. It was to predict that will new customers will become a lead and generate the company's expected profit.

Step 2: Analysis, Modelling, and Validation

I have used target variable as **"Avg_Sale_Amount"** while the predictors include **"Avg_Num_Products_Purchased"** and **"Customer Segment".**

Table | Report | Profile
1 of 1 Fields | Records 1 to 10

Record	Report
1	**Report for Linear Model Linear_Regression_Sales**
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		

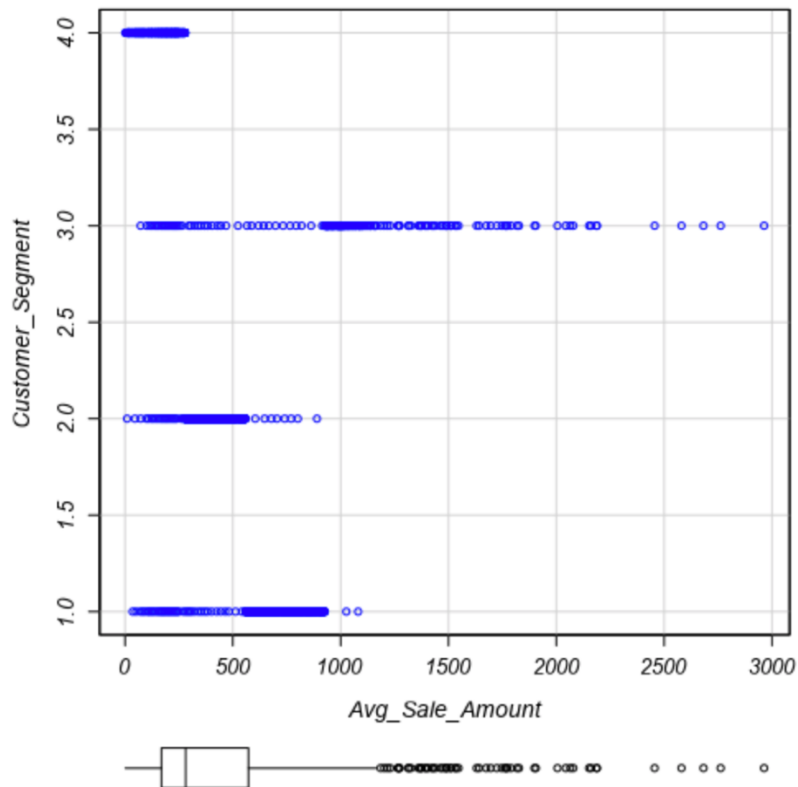Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

These predictors were chosen because of their **p-values** which is far less than 0.05 and **Adjusted-R Squared** is 0.8366.

Relation between Avg_Sales_Amount and Avg_Num_Prodcuts Purchased
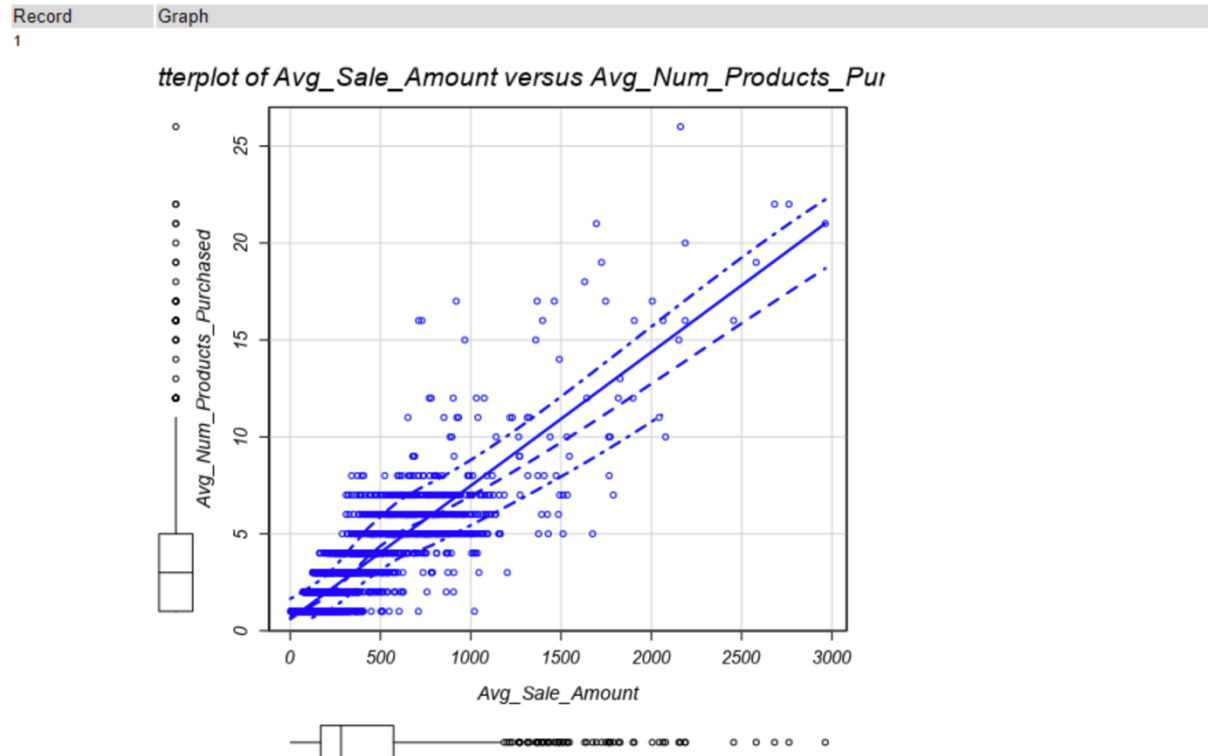
Relation between Avg_Sales_Amount and Customer Segment

Scatterplot of Avg_Sale_Amount versus Customer_Segment

Linear Regression Equation

Expected_Avg_Sale_Amount= 303.46 + (-149.36)*(Customer Segment: Loyalty Club Only) + (281.84)*(Customer Segment: Loyalty Club and Credit Card Only) + (-245.42)*(Customer Segment: Store Mailing List) + (66.98)*(Avg_Num_Products_Purchased)

Step 3: Presentation/Visualization

What is your recommendation?

The company should send these 250 catalogs to the new customers.

How did I come up with the recommendation?

Predicted_Average_Sales= Expected_Avg_Sale_Amount * Score_Yes

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

Profit= (Predicted_Average_Sales * 0.5) – (Cost of Catalog * 250)

Where **Cost of Catalog is $6.50**

The profit came out to be $21,987.435

Record	Sum_Expected_Average_Sales	Profit
1	47224.871373	21987.4356865455

Alteryx Workflow: