Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

To predict the expected profit from the 250 new customers and whether to send or not the catalog to the new customers based on expected profit.

2. What data is needed to inform those decisions?

The following are data needed to predict sales and calculate expected profit:

- p1-customers.xlsx: The dataset that includes information on about 2,300 existing customers.
    - Customer_Segment
    - Avg_Num_Products_Purchased

- p1-mailinglist.xlsx: The dataset on the 250 new customers that is required to predict sales.
    - Avg_Num_Products_Purchased
    - Score_Yes

- Gross Margin
- Cost of Catalog

Analysis, Modeling, and Validation

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

Steps Taken:

1. A linear regression study is performed on the file p1-customers.xlsx.

2. 'Customer_Segment' and 'Avg_Num_Products_Purchased' were chosen as predictor variables. 'Responded_to_Last_Catalog' not chosen as it is not relevant to new customers we want to learn about and the rest of the variables are address which have no impact on new purchases.

3. 'Avg_Sale_Amount' is selected as target variable.

Report for Linear Model Linear_Regression_7

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

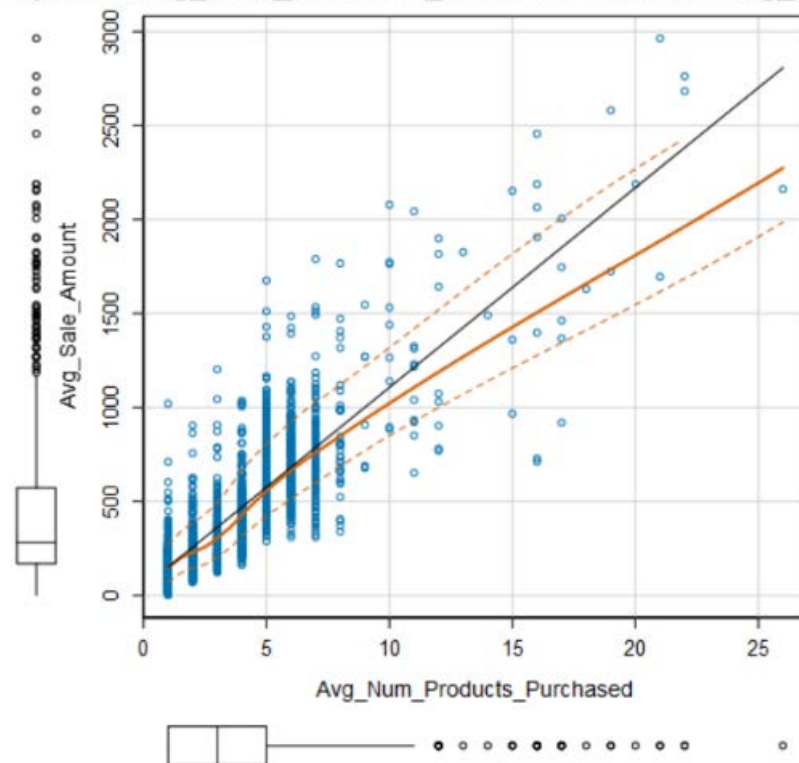*Figure 1 Linear Regression Report*

Figure 2 Avg Num Products Purchased vs Avg Sale Amount Scatterplot diagram

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

The Alteryx linear regression function is used to determine the strength of the linear and the statistical result shows an adjusted R-squared value of 0.8366 which is a high value. We can also see that both Customer Segment and Average Number of Products Purchased have a p-value of less than 0.05 which means we can be 95% confident that there exists a relationship between the predictor and target variable. Thus, the model is considered a good one.

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

Y = 303.46 + (281.84 X Customer_SegmentLoyalty Club and Credit Card) + (-149.36 X Customer_SegmentLoyalty Club Card Only) + (-245.42 X Customer_SegmentStore Mailing List) + (66.98 * Avg_Num_Products_Purchased) + Credit Card * 0

Presentation/Visualization

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

The Management does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000. The expected profit is predicted by creating the linear regression model using the past customers and fitting the new 250 customer's data on this model. Since the predicted profit contribution exceeds $10,000, we can recommend the company to send out the catalog to the new customers.

2. **How did you come up with your recommendation?**

1) The expected sale from each customer is determined by multiplying the **Score** field (sales amount) with **Score_Yes** value.
2) With a gross margin of 50%, 50% is deducted from the sum of **Predicted_Sale** before the cost of catalog ($6.50) is subtracted to obtain net profit.

3. **What is the expected profit from the new catalog?**

Expected Profit = *(Sum of predicted sale x Gross Margin) - (Cost of Catalog x 250)*

The expected profit we can expect from the new catalog assuming it is sent to the 250 customers is **$21,987.44**.
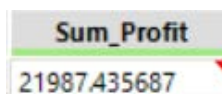
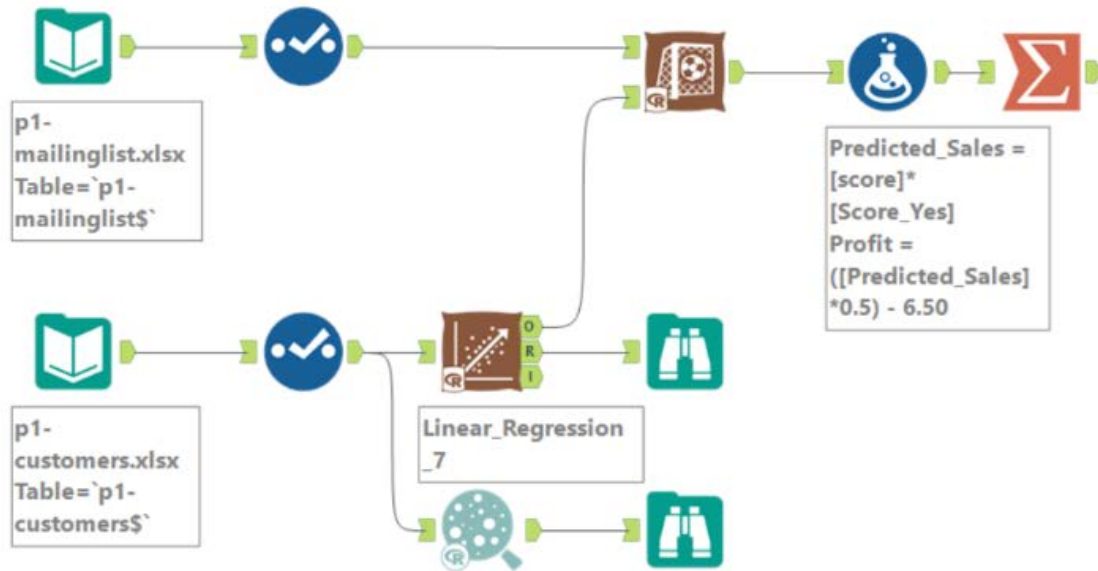| Sum_Profit |
|---|
| 21987.435687 |

Figure 3 Sum of Profit

p1-
mailinglist.xlsx
Table=`p1-
mailinglist$`

Predicted_Sales =
[score]*
[Score_Yes]
Profit =
([Predicted_Sales]
*0.5) - 6.50

p1-
customers.xlsx
Table=`p1-
customers$`

Linear_Regression
_7

Figure 4 Alteryx Flow