Project 1: Predicting Catalogue Demand

Business Problem:

A company manufactures and sells high-end home goods. Last year the company sent out its first print catalogue, and is preparing to send out this year's catalogue in the coming months. The company has 250 new customers from their mailing list that they want to send the catalogue to.

Your manager has been asked to determine how much profit the company can expect from sending a catalogue to these customers. You, the business analyst, are assigned to help your manager run the numbers. While fairly knowledgeable about data analysis, your manager is not very familiar with predictive models.

You've been asked to predict the expected profit from these 250 new customers. Management does not want to send the catalogue out to these new customers unless the expected profit contribution exceeds $10,000.

What decision needs to be made?

The decision here is that the company wants to launch a new catalogue for their new high-end home goods. The costs of printing and distributions is $6.50 per catalogue. In order to get the expected profit of more than $10,000, we need to analyse each metrics and factors that contribute to the sales of the company.

What data is needed to inform those decisions?

The data needed will be:

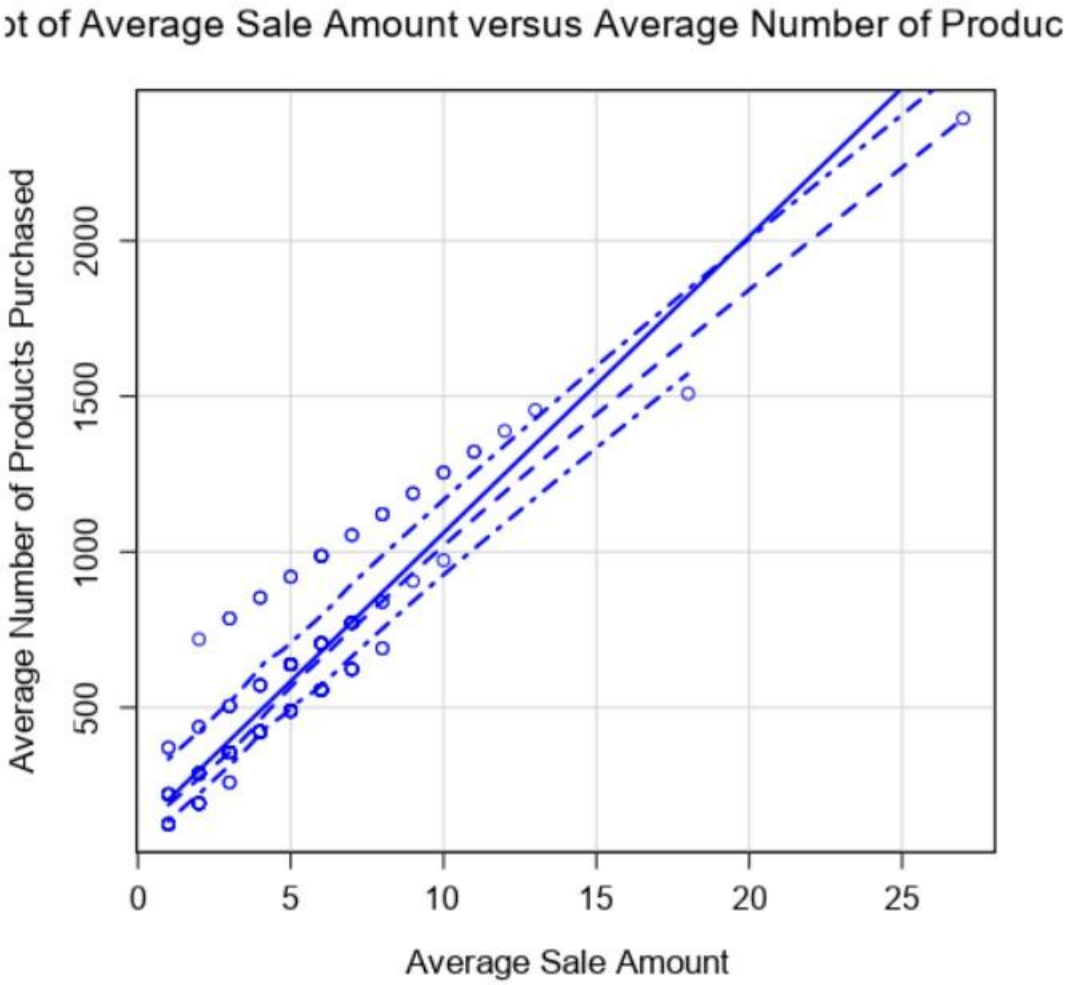
- *Customer_Segment*
- *Store_Number*
- *Avg_Num_Products_Purchased*
- *Years_as_Customers*
- *Avg_Sale _Amount*

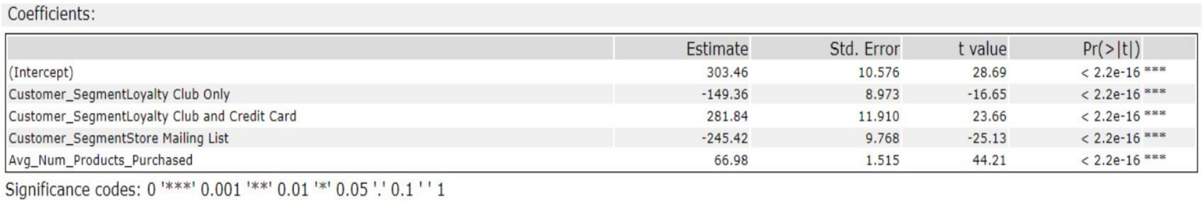
Analysis, Modelling and Validation

The target variable here is the average amount of sales by a customer. In our model, the predictor variables chosen are:

- *Customer_Segement*
- *Avg_Num_Products_Purchased*

If we plot a scatterplot between the average number of products purchased on the X axis and the average sale amount on the Y-axis, we can observe the linear increasing trend in the graph.



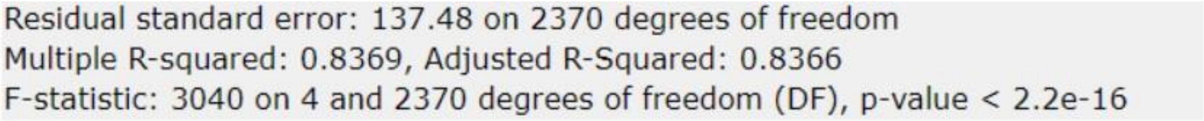

Coefficients:

	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

If we consider the p-value of all the predictor variables, the value is less than 2.2e-16 which is a low value, thus there exists a relationship between the predictor and target variable. The p-value tests the null hypothesis that the coefficients is equal to zero. Hence with a low p-value, the null hypothesis can be rejected.

Customer Segment is a categorical variable in our data. From the above report, the p-value generated from our model shows that for each value in the Customer Segment, there is a very small value of p (<2.2e-16) indicating a great relationship between the target variable and the predicted variable making it a significant variable.



```
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
```

The Adjusted R-Squared value for our model is 0.8366. This high value tells that the model good in predicting the prices and is able to explain a lot of variation in the prices.

From creating the model, we get the linear regression equation as:

Y = 303.46-149.36(Customer_SegmentLoyalty Club) + 281.4*(Customer_SegmentLoyalty Club and Credit Card) - 245.42*(Customer_SegmentaStore Mailing List) + 66.98*(Avg_Num_Products_Purchased)*

Presentation/Visualization:

The expected profit from the new catalogue after sending the it the 250 customers is $22800 which surpasses the $ 10,000 margin that the company were expecting.

I recommend that the company should send the catalogue to the 250 customers. It will be beneficial for the company.

After modelling the linear regression with the data of the customers, we get the predicted amount. The data contains the ‘Score_Yes’ figures which gives the probability that the customer buys the product. Hence

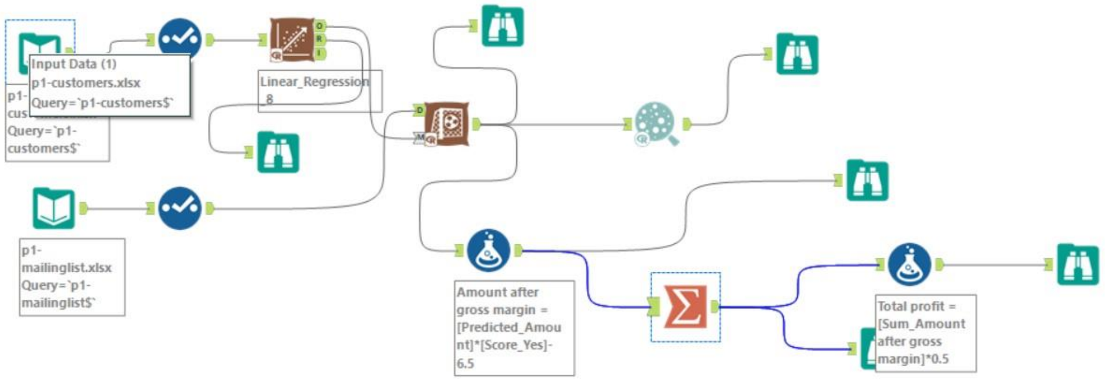
Predicted amount= Amount * Score_Yes

Since the company spends $ 6.50 for the printing and distribution of each catalogue, we need to deduct the number to the predicted amount.

Predicted Amount= Predicted Amount- 6.5

Since the average gross margin on all the products sold through the catalogue is 50%. Hence, we need to multiply the total predicted amount with 0.5 to get the profit.

Expected Profit= Total of Predicted Amount * 0.5



The above photo is the Alteryx workflow where the model was created, the prices were predicted and the expected profit was calculated.