# *Predicting Catalog Demand*

# Step 1: Business and Data Understanding

The company sells high-end home goods, Last year the company sent out its first print of the catalogue. In the coming months they we will prepare the catalogues for this year send. Management wants to determine if they should send out 250 catalogues to the new customers. If the expected profit exceeds $10,000 then the catalogues will be sent to this new customer group

## Key Decisions

Key Decisions: Is the expected profit >10,000, if case is true then the company will send the catalogues.

Data required to be able to determine the decision is as follows:

* The cost of printing & distribution per catalogue, $6.50.
* Average gross margin (price-cost) on all products sold through the catalogue is 50%.
* Multiply revenue by gross margin first before subtraction the costs.

Existing & new customer data that includes:

* Average sales amount per customer.
* Customer response to catalogue last year.
* Customer payment method.

We are a data rich company and the existing and new customer details have been provided in two files.

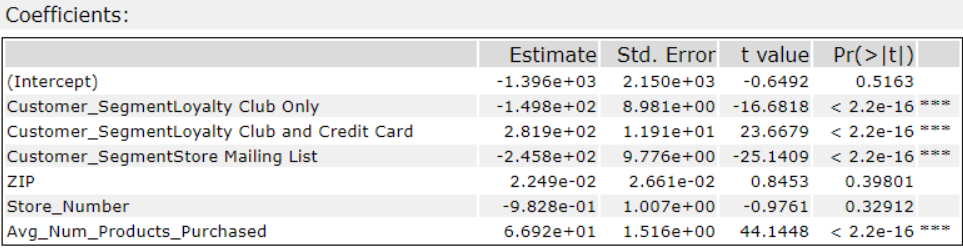
# Step 2: Analysis, Modeling, and Validation

We will exclude the following variables:

* Name: sales doesn’t depend on a name, and we can get many false results due to duplicated names
* Customer\_ID: is assigned to the name and is a unique ID, not having impact on avg\_sales
* Address: Also unique to a customer and will not have an impact on avg\_sales.

I will explore the data and use the visual exploration method, to check what relationships there are between the target variable: average sale amount.

I decided to look at few numeric variables for the predictor, at *post\_code* and *store\_number and also customer segment* as these are present in both sets of data.



Looking at the statistical results there is significant coefficients between *Customer\_Segment* and A*vg\_Num\_Products\_Purchased* as these are <0.05. We also see a this with *ZIP* and *Store\_Number* is not statistically significant as the p\_value >0.05.

Plot of *ZIP* and *Store\_number* to visualize the data.

|  |  |
| --- | --- |
|  |  |

Looking at the plots we see there is no linear relationship as the slope is close to zero when comparing the variable for *Avg\_Sale\_Amount*.

Plot Avg\_Number\_Products\_Purchased

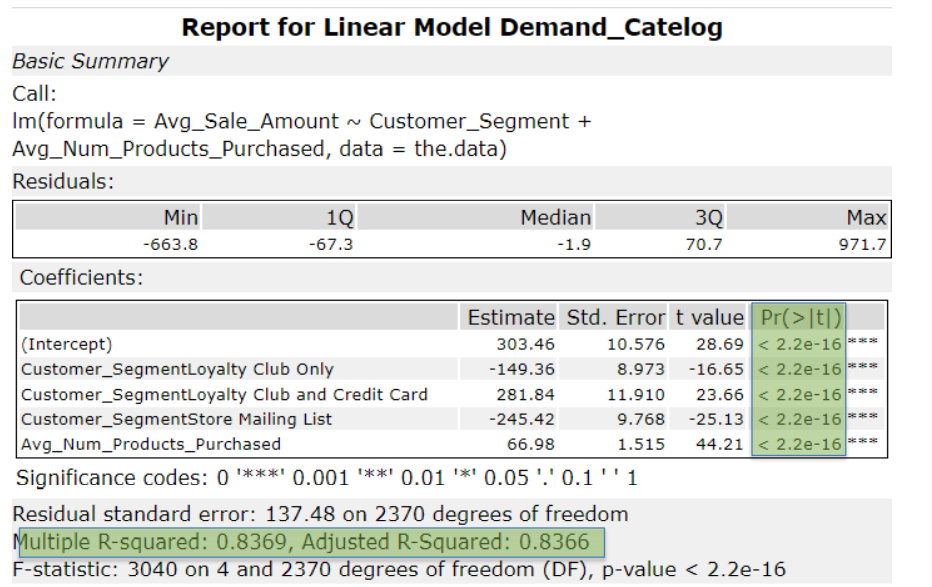
|  |  |
| --- | --- |
|  |  |

Using the average number of products and average sales we see that there is a strong linear trend The correlation matrix shows we can further see that the Avg\_Num\_Products\_Purchased the Avg\_Sales\_Sale is strong with an R value of 0.86. The Scatterplot shows that the is a linear relationship between the two. This is reflected in the Statisical summary showing the coefficients.

We can build our model using the Customer\_Segment, Avg\_Number\_Products\_Purchased.

# Step 3: Presentation/Visualization

The statistical results of the model.



The statistical result above show that the regression model is good as each predictor variable has a P value > 0.05. The second reasoning that this is a good model, is that the Adjusted R-Squared value is close to 1, with a value of 0.8366. This suggest that the predictor variables and the target variable in the model are statistically significant.

The best linear regression equation is as follows.

Avg\_Sale\_Amount = 303.46 + 66.98 \* Avg\_Num\_Products\_Purchased -149.36 (If Customer\_Segment: Loyalty Club Only) + 281.84 (If Customer\_Segment is Loyalty Club and Credit Card) – 245.42 (If Customer\_Segment is Store Mailing List) + 0 (If Customer\_Segment is Credit Card Only)

## Applying Business Rules

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

My Recommendation is that the company should sent the catalogs to the 250 new Customers, as the predicted profit is greater than the stated $10,000.

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

After applying the model to mailing list, to get the predicted\_sales\_amount we applied the following as per requirement.

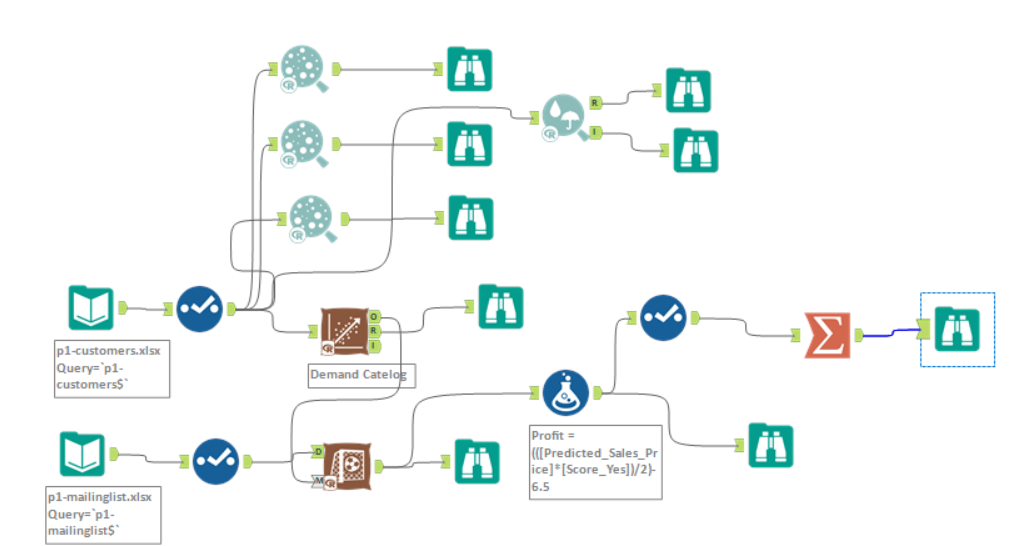
Multiply the predicted\_sale\_amount by score\_yes (Score\_yes is the likely hood the customer will buy) for each customer. Sum the amount above and multiply by the 50% gross margin. Then subtract $6.50 per customer.

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

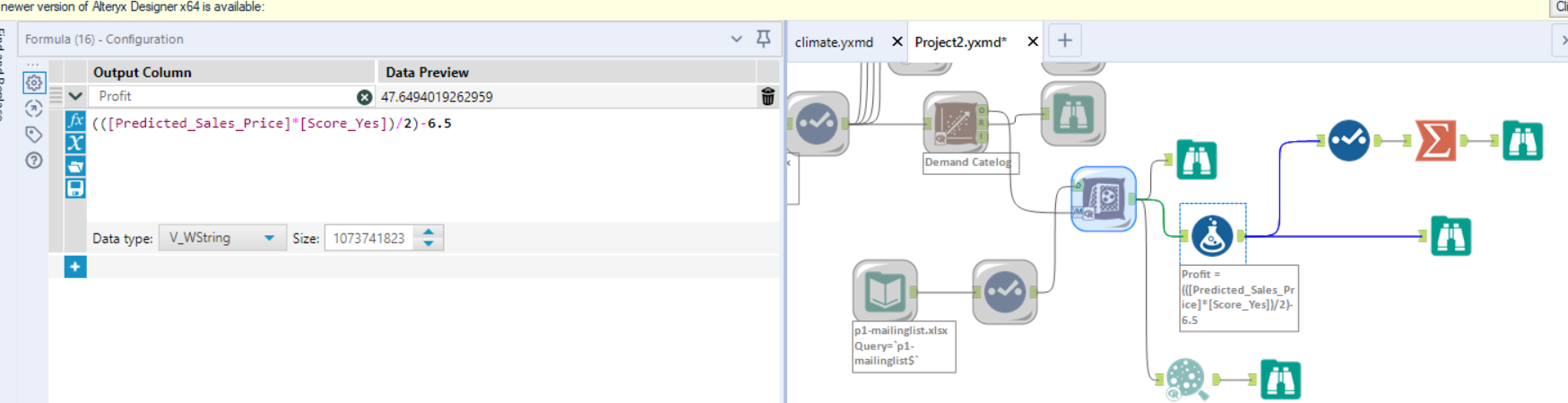
$21,987.44 = (sum([Predicted\_Sale\_Amout]\*[Score\_Yes])\*.5)-6.5

# Alteryx Workflow

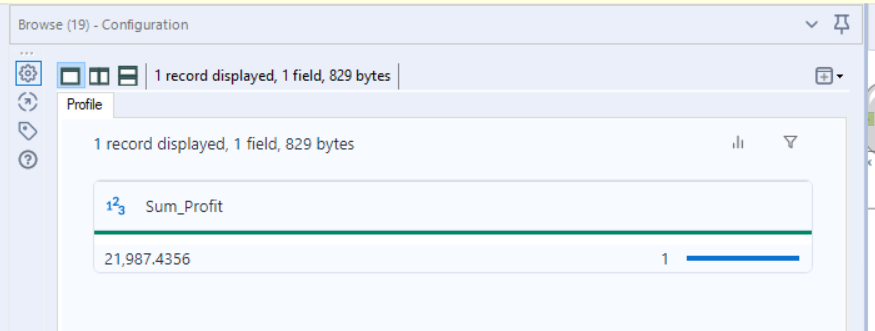
### Workflow



### Applying Business Rules



### Calculated Profit



# Resources

<https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/>

<https://courses.lumenlearning.com/wm-macroeconomics/chapter/interpreting-slope/>

Supporting Model in Excel, to check usage of Alteryx