Project 1: Predicting Catalog Demand

## **Business and Data Understanding**

The main outcome is to predict whether a new catalogue release will generate enough profit to justify releasing the catalogue. We need to predict the amount of sales the catalogue would generate when it is sent out to the company’s mailing list. Moreover, the company needs to make a 50% margin and to also generate sales greater than the business’ own requirement of $10,000.

**What decisions needs to be made?**

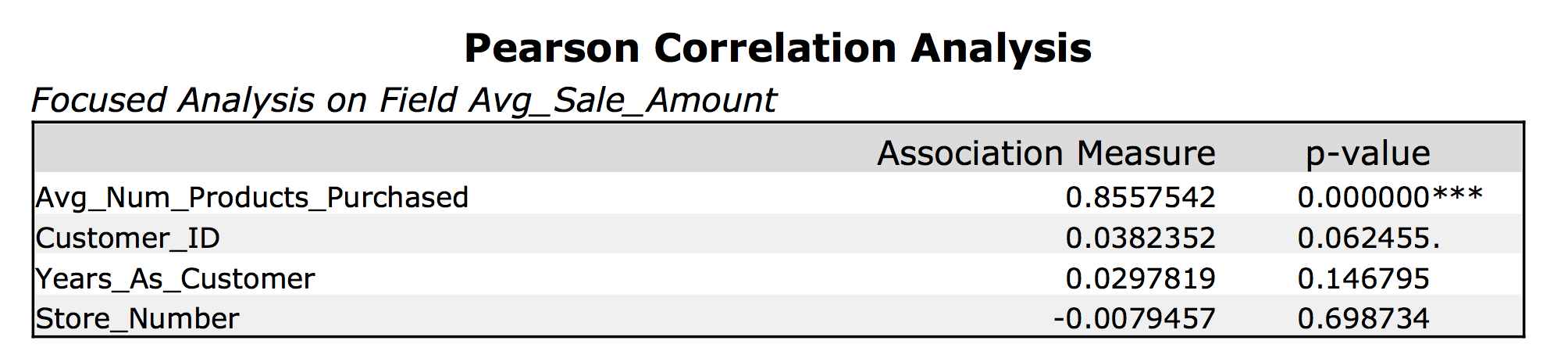
The decision here is to decide whether to go with a new catalogue launch for their high-end home goods product range. There is a cost to producing these catalogues and an expected profit margin. The sum of the profits after costs and margin needs to also be greater than $10,000 in order for this catalogue to be launched successfully.

**What data is needed to inform those decisions?**

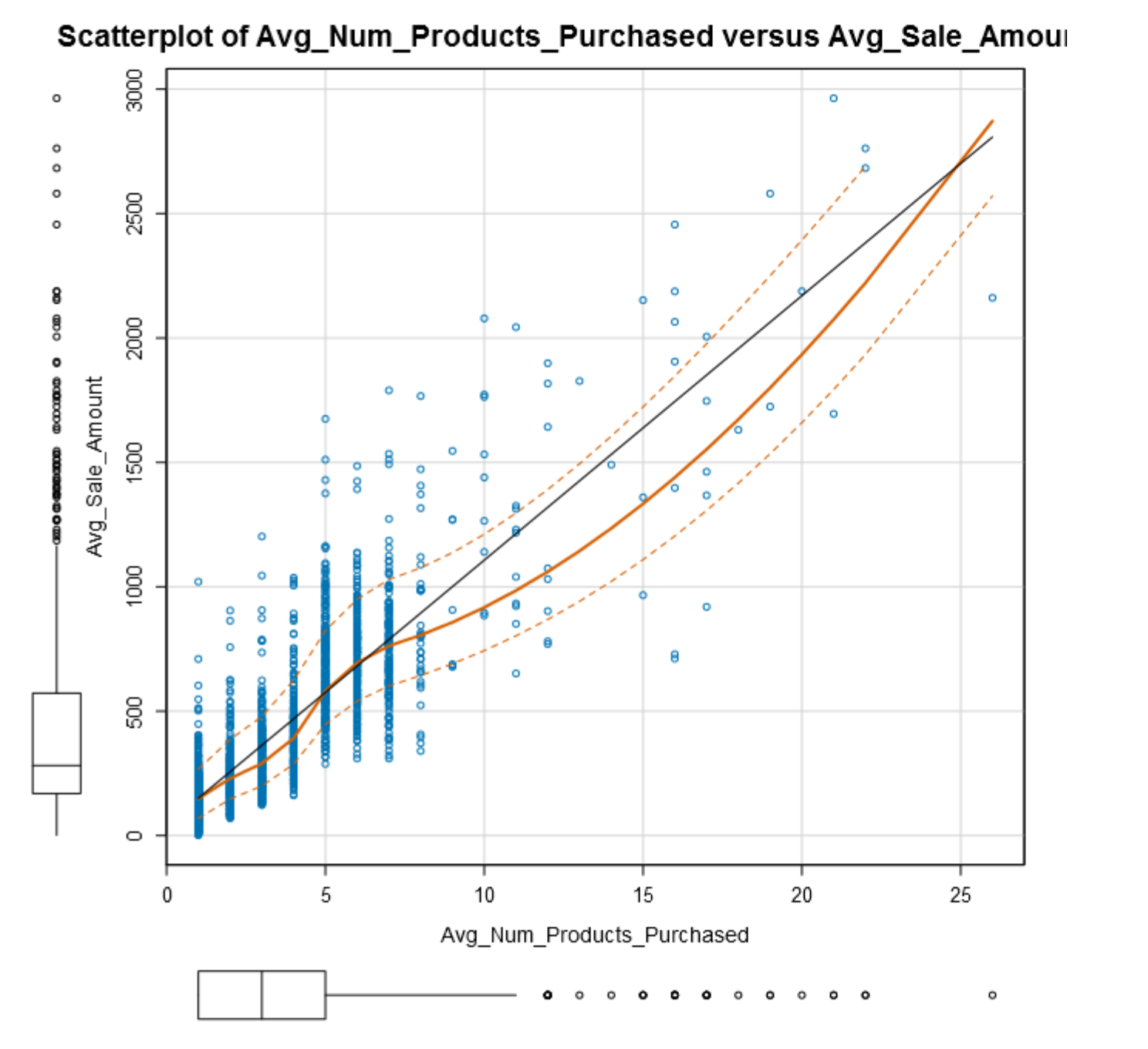
* Average sales amount
* Average number of products purchased
* Number of years as customer
* Store number
* Customer segment

## **Analysis, Modeling, and Validation**

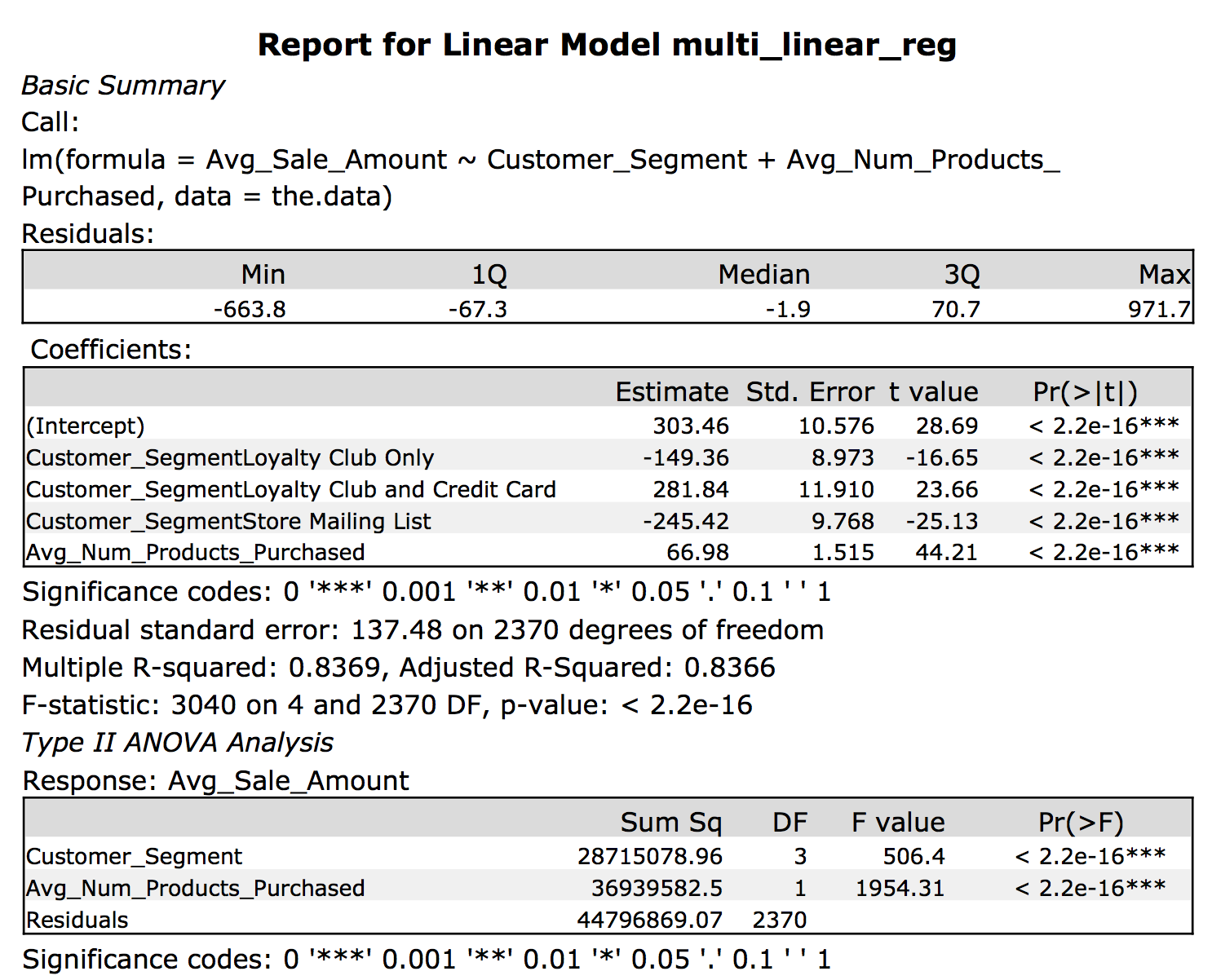
The Pearson’s Correlation between all the predictor variables against the target variable (Average Sales Amount):



Clearly, Avg\_Num\_Products\_Purchased is the best variable with the highest association measure, p-value of 0, and with three stars.



Customer Segment is a categorical variable found in our data and a variable that needs to be tested to see if there is any correlation with the target variable. Using Alteryx and running a linear regression using Average Number of Products Purchased and Customer Segment, we get the below summary:



Evidently, both variables are highly useful in predicting the average sales amount as shown by the adjusted R-Squared value and the three stars across the board.

Therefore, the linear equation is:

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

## **Presentation/Visualization**

## To obtain the actual profit, for each customer, we first multiply the predicted sales with the probability of the customer buying the products (the “Score\_Yes” variable), and then multiply that value with 0.5 (which accounts for the margin), and then finally subtract $6.50. After summing all the individual profits, we get the total profit as $21,987.44.

Since this value is greater than $10,000 that was the baseline for the company, I would recommend that the company go ahead with the catalogue launch. This recommendation can be trusted as the model built above is very robust and accurate.