Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

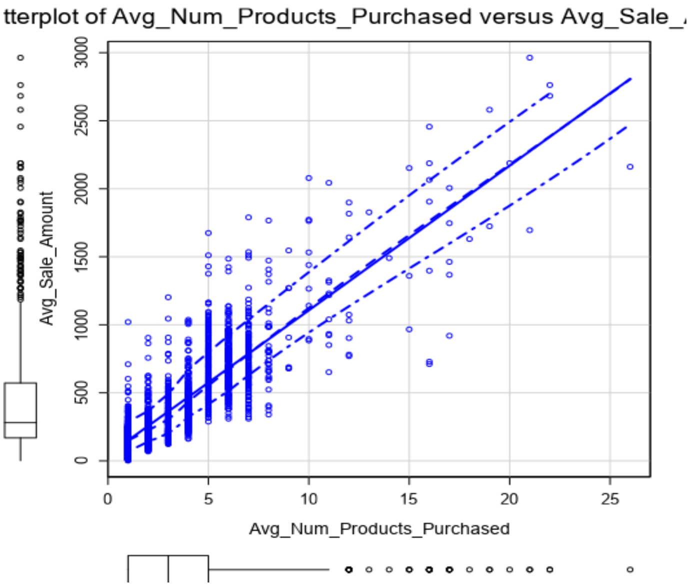
A company that manufactures and sells high-end home goods *needs to predict whether or not it will be profittable to send out this year's catalog to a list of new cutomers.*
The cost for printing and distributing the new catalog is $6.50 each and the goal for the company is to make a profit of at least $10,000.
For the prediction the company will provide the list of the old customers with their response to the last year's catalog campaign.

Key Decisions:

We need to decide whether or not to advice the company to send out the catalog to these new clients based on the fact that the predicted profit caused by the campaign will be enough or not.
In order to make this decision we need to look into the last year campaign's data and create a regression model that analyzes the data of the old customers and predict the profit that sending the catalog to a new customer would bring.
It is important to understand what data will be usefull for the model and what will not be, for this model the best predictors will be the customer's segment and the average number of product purchased.

Step 2: Analysis, Modeling, and Validation

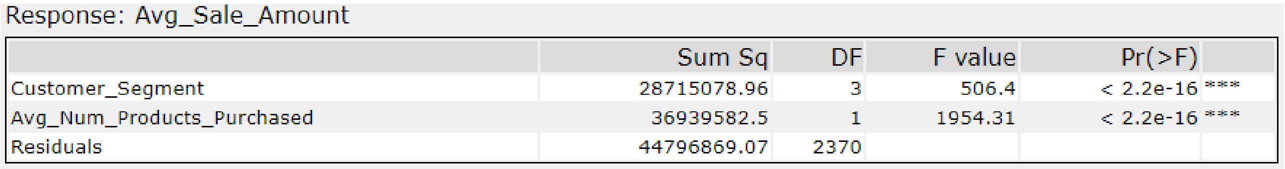
The variables that I have choosen to use are "cutomer segment" and "average number of products purchased" because their P-value suggested that they were the most relevant for the regression



The scatter plot between the average sale amount and th average number of products purchased show how the varable influence the target.

The Adjusted R-squared value of the model is 0.8366 so it should make predictions that are 83% right.

I have also tried to insert some categorical variables like "City" or "State" but their p-value was to low and the R-score did not improved by using those variables. The same thing apppend if you use the variable "ZIP code".



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			

Average Sale Amount= 303,46 -149.36 * Customer_Segment *(*Loyalty Club Only *)* +281.84 * Customer_Segment *(*Loyalty Club and Credit Card) -245.42 * Customer_Segment *(*Store Mailing List *)* + 66.98 * Avg_Num_Products_Purchased

Step 3: Presentation/Visualization

The model, analyzing the data of the old sutomers predicted a profit of $21,987.44, way over the treshold of $10,000 so the recommendation for the company is to send out the new catalogs to the new customers.

I have decided to use only two variables for this model because the model were perfoming better with those two variables. After the model made the predictions for the new clients I have moltiplied those to the probability that the new customer would purchase something from the catalog (Score_yes). After that I have subtracted the catalog's cost from every customer's expected profit and then I have summed the expected prof for every sutomer and obtained the prediction for the campaign profit.