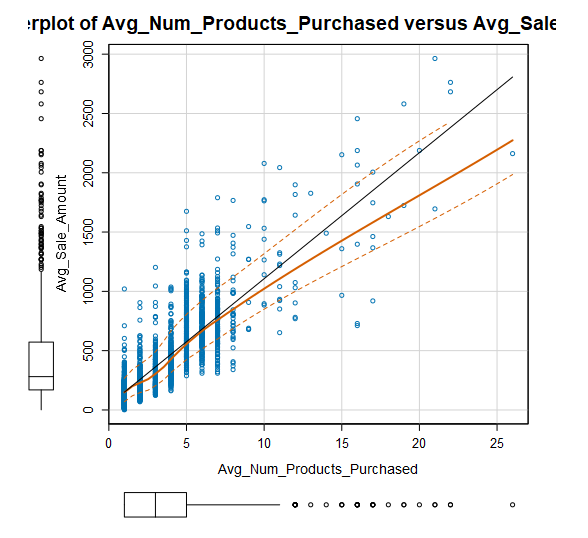
Project 1: Predicting Catalog Demand

## **Step 1: Business and Data Understanding**

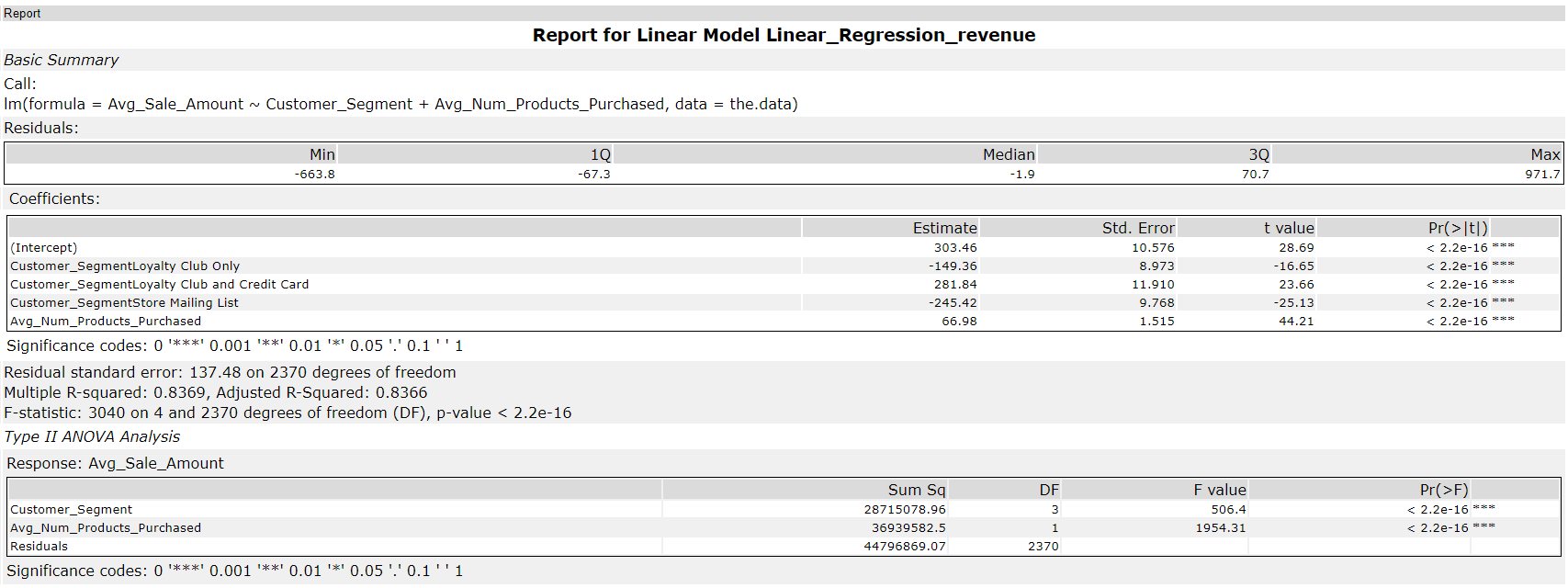
***The company has 250 new customers from the mailing list. We are interested to send them the catalog and get extra business. We used existing customer data to predict the expected profit from these 250 new customers if we sending them the catalog. We are only interested to send them the catalog if the expected profit exceeds $10,000.***

## **Step 2: Analysis, Modeling, and Validation**

***Existing customer dataset have both numeric and categorical data.***

***For numeric data, we used the scatterplots to see if the variable is a good candidate for a predictor variable. I tried ZIP code, Store Number, Year as customer and Avg number products purchased. And it turns out that only avg number of products purchased have a positive relationship with sales amount.***

***For categorical data, we used trial and error to see if they are statistically significant. I checked P value on each variable and to see if they have a strong relationship with Sales amount. It turns out that customer segment is an important categorical variable to predict sales amount.***



***After we build our model, we find our adjusted R-squared is 0.8366 and all P value are below 0.05. A high R-squared and low P-values indicate it’s a good model.***

***The linear regression equation is below:***

***Ave sales amount = 303.46 + 66.98 (Avg\_Num\_Products\_Purchased) – 149.36 (If Customer Segment : Loyalty Club only) + 281.84 ( If Customer Segment : Loyalty Club and Credit Card) - 245.42 (If Customer Segment : Store Mailing List)***

## **Step 3: Presentation/Visualization**

I will recommend to send out the catalog to these 250 new customers. Based on our analysis, we will have an estimated profit for $ 21,987, which exceeds $10,000.

***The way we calculate the estimated profit is below:***

***Expected profit for each new customer = Predicted avg sales amount \* Score\_Yes \* 0.5 - 6.5***

***Total Expected Profit = Sum (expected profit for each customer) = $ 21, 987***