Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

1. **What decisions needs to be made?**

   We need to determine how much profit the company can expect from sending a catalog to these 250 new customers. So, we need to analyze the company's previous data to build prediction model. If we are confident in our prediction model results, we will run the mailing list data, and predict the sales amount and potential customer profit. Then sending the catalog out to these new customers if the expected profit contribution exceeds $10,000.

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

   There will be many previous and historical data needed to be analysing to predict the maximum benefit of catalog sending process. We need to completed data about **customer segment, average number of products purchased, the average amount of sales to build prediction model**. After the predictive process many data are produced, these are **the score**, **the predicted sales, and the profit of these sales.**

   And to get the final profit values we need to use some information that mentioned in project details:

- The **costs** of printing and distributing is **$6.50** per catalog.
- The average **gross margin** (price - cost) on all products sold through the catalog is **50%**.
- And multiply the **revenue** by the gross margin first then subtract out the $6.50 cost to calculate the **profit.**

Step 2: Analysis, Modeling, and Validation

1. **How and why did you select the predictor variables in your model?**

I choose numeric variable with categorical variable (Avg_Num_Products_Purchased, and Customer Segment) as predictors variables. They both have positive relationship with targeted variable(Avg_Sales_Amount). These variables have a close linear relationship, the increments or decrements of one will effects on another.

The scatter plot below represents the relationship between them:

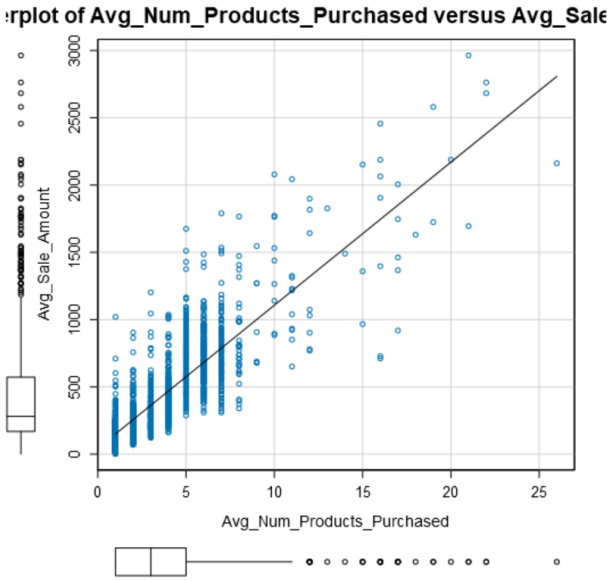
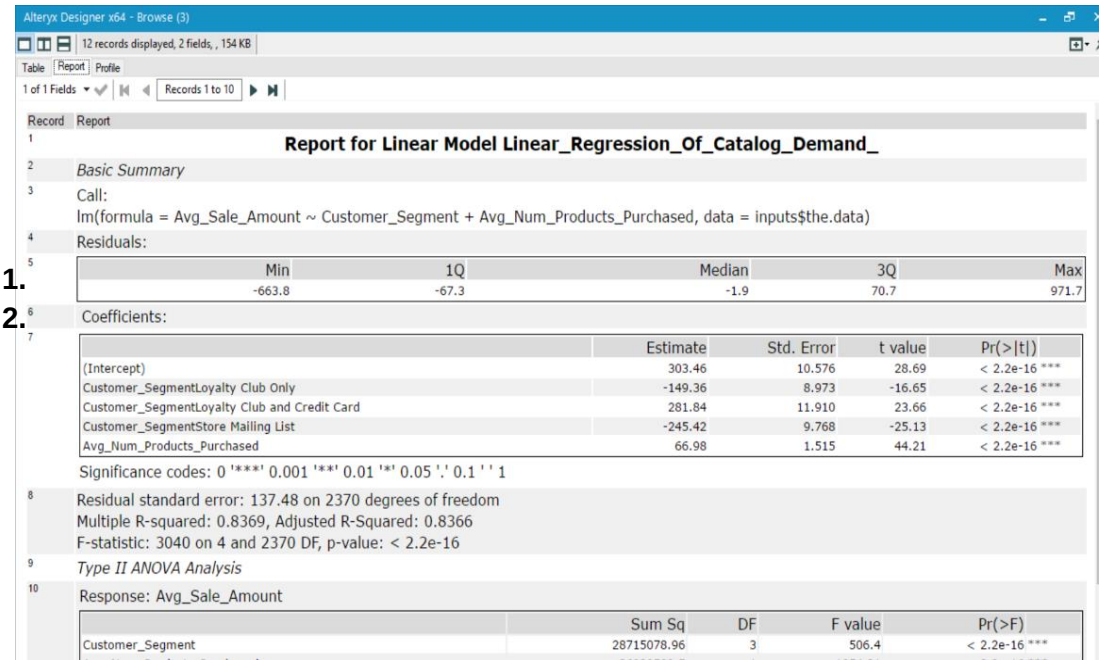


Figure 1:scatter plot of average number of products purchased, and average sales amount

If there is a slope to the line, then this might indicate that this is a good predictor variable for this target variable. The graphs above would indicate the (Avg_Num_Products_Purchased, and Customer Segment) would be good candidates to be predictor variables for the target variable (Avg_Sales_Amount).



1.
2.

Figure 2: Report for Linear Model

3. **Explain why you believe your linear model is a good model. You must justify your reasoning**

   I have built a linear model with strong linear relationship between the variables. The slope represents the positive increase. And as we know low P-values (<0.05) and a high R-squared (>0.7) suggest the model is highly predictive. I get the R-square value with 0.8396 (above the 0.7 threshold), and the model predicts the coefficients with P-values less than 0.05.

   When we look out to the right to the Significance codes. They have one or more asterisks, which means it is statistically significant. If one of the individual categories is statistically significant it is a good categorical variable to consider for the model. So, I have chosen Customer Segment as a categorical variable (with ***).

3. **What is the best linear regression equation based on the available data?**

   The best linear regression equation is:

   Y = 303.46 - 149.36(If Loyalty Club only) + 281.84 (If Loyalty Club only & credit card) - 245.42 (If Store mailing list) + 66.98 (Avg_Num_Products_Purchased) + 0

   Note: 0 in the equation represents the Credit Card Only predicted coefficient

Step 3: Presentation/Visualization

answer these questions:

1. **What is your recommendation? Should the company send the catalog to these 250 customers?**

   I recommend to the company to send the catalog to these 250 new customers for many reasons from my point of view:

   - First, the predicted profit value is high.
   - Second, the cost of catalog for 250 customers will be covered by the final profit values 13 times.
   - There will be no significant financial consequences because all costs will be covered by revenues.

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

   - After linear regression model built, I used the linear regression equation to predict sales for each new customer.
   - aggregated the sales data and multiplied the total by 50% (per management's average gross margin (price - cost) on all products sold through the catalog) to get the expected profit.
   - Finally, I subtracted the cost of mailing catalogs to the new 250 customers, which gave me the final predicted profit for these new customers.

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

   The expected profit is $21987.435.