Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

decide whether to send the printed catalog to the new customers or not, decide so depending on analyzing old customers' response to sending the catalog to them last year.

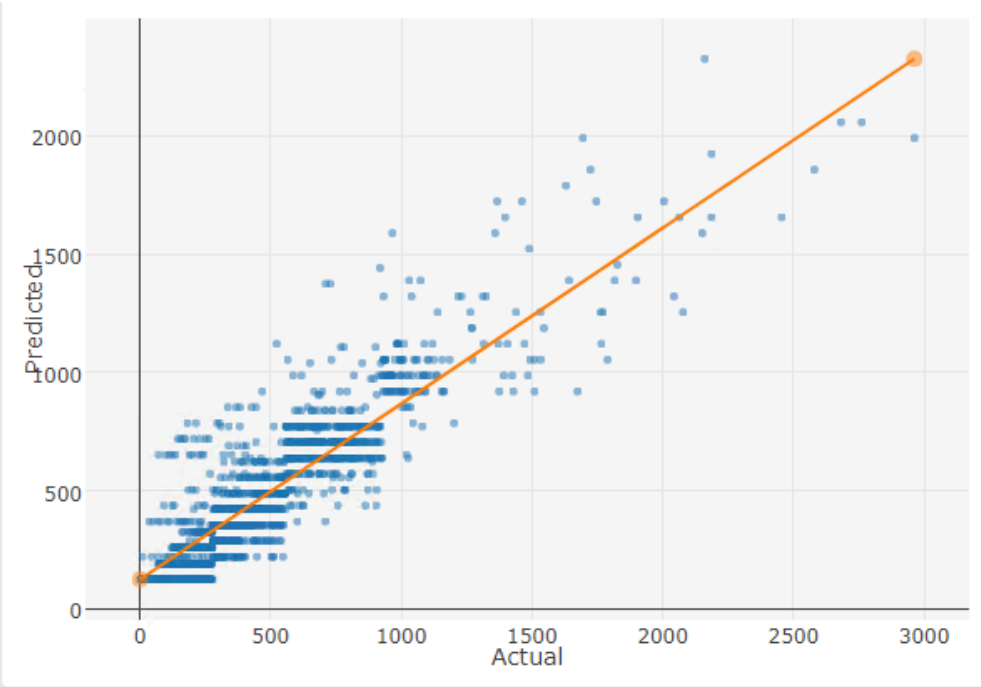
2. What data is needed to inform those decisions?

Data of old customers and the new ones, which includes customer segment and average number of products purchased.

Step 2: Analysis, Modeling, and Validation

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

Customer_segment and Avg_Num_Products_Purchased have linear relationship with average sales amount, depending on which customer segment and how many products are purchased we can predict sales amount. (Scatter plot should be attached)



2. Explain why you believe your linear model is a good model.
By using customer_segment & avg_number_products_purchased as predictors, we have P-value of 2.2e-16 for all predictors and R-squared equals to 0.8369 which indicates a good model used here.

3. What is the best linear regression equation based on the available data?
Y = 303.46 + 66.98*Avg_Num_Products_Purchased – 149.36*(if loyalty club only) + 281.84*(if loyalty club and credit card) – 245.42*(if store mailing list)

Step 3: Presentation/Visualization

1. What is your recommendation? Should the company send the catalog to these 250 customers?
Yes they should send the catalog to the new customers.

2. How did you come up with your recommendation? (Please explain your process so reviewers can give you feedback on your process)
The expected total revenue is $21987.46 (which exceeds the $10000 threshold), so it's worth spending $1625 as sending catalog cost.

3. What is the expected profit from the new catalog (assuming the catalog is sent to these 250 customers)?
The expected total revenue is $21987.46.