<u>Project 1: Predicting Catalog Demand</u>

Step 1: Business and Data Understanding

Provide an explanation of the key decisions that need to be made. (500 word limit)

Key Decisions:

Answer these questions

1. What decisions needs to be made?

    - *We need to build prediction model for a new 250 customers, and see if it is going to be profitable for us to send them a catalogue.*

2. What data is needed to inform those decisions?

    - *We need mailing list of new customers, percentage probability with a positive outcome, margins and expenses of catalogue creation and distribution.*

Step 2: Analysis, Modeling, and Validation

Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500 word limit)

Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

1. How and why did you select the predictor variables in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer back to the "Multiple Linear Regression with Excel" lesson to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.

- *I have selected Customer_Segment and Avg_Num_Products_Purchased predictors to perform Linear regression model, with whom I am going to train a prediction model and get a target variable (average sales amount) of a new 250 customers. I have used scatterplot to see the correlation of those two variables, and noted that there is a good correlation between them.*

tterplot of Avg_Sale_Amount versus Avg_Num_Products_Pur
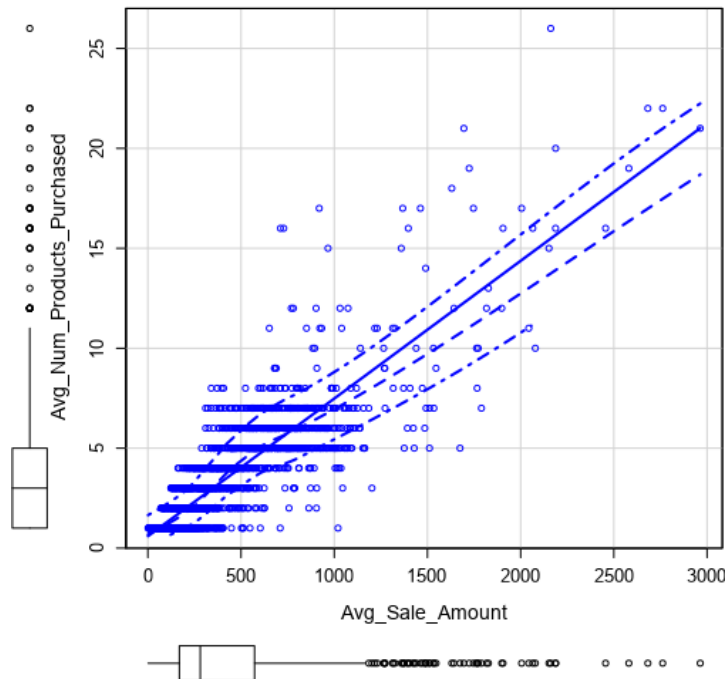
2. Explain why you believe your linear model is a good model. You must justify your reasoning using the statistical results that your regression model created. For each variable you selected, please justify how each variable is a good fit for your model by using the p-values and R-squared values that your model produced.

- *I have concluded that I can use obtained prediction model because I have statistical significance between variables. My R squared, and Adjusted R squared are high 0,83. Also the p value for all predicators are way below 0,05.*

3. What is the best linear regression equation based on the available data? Each coefficient should have no more than 2 digits after the decimal (ex: 1.28)

- *$Y = 303.46 - 149.36 * Customer\_SegmentLoyalty$ Club Only $+ 281.84 * Customer\_SegmentLoyalty$ Club and Credit Card $- 245.42 * Customer\_SegmentStore$ Mailing List $+ 66.98 * Avg\_Num\_Products\_Purchased$*

Step 3: Presentation/Visualization

Use your model results to provide a recommendation. (500 word limit)

At the minimum, answer these questions:

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

- *Yes. The company should send the catalogue to new customers, because the estimated profit is 21987.43$, above low limit of 10000$.*

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

- *After I have obtained linear regression result model, I had to combined it with a data of new customers, so I could be able to calculate Average sales amount per customer. Then I have used three formulas. Firstly, I have multiplied the result with a probability of buying a product per person. Secondly, I have multiplied given result by the average gross margin (0.5). At the end I have subtract it with the costs of printing and distributing of catalogue (6.5$ per catalogue), and summed up to get a final expected profit.*

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

- *The expected profit is 21987.43$.*