Project 1: Predicting Catalog Demand

## **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?

* The decision manger needs to make is " Can we expect profit more than 10 thousand if we send the catalog to the 250 customers?"

1. What data is needed to inform those decisions?

* We need Average sales per customer so we can calculate the revenue then multiply it by gross margin and then subtract it by 6.5 to finally get the revenue .once we got the sum of profit to all 250 customer in the mail list we will see if it more than 10000 we will send the catalog otherwise we will not.

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

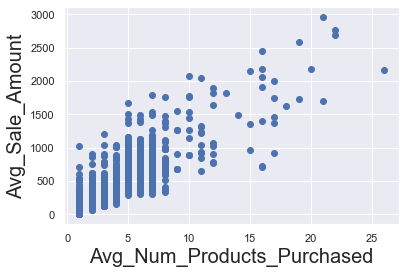
*Provide a description of how you set up your linear regression model, what variables you used and why, and the results of the model. Visualizations are encouraged. (500-word limit)*

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

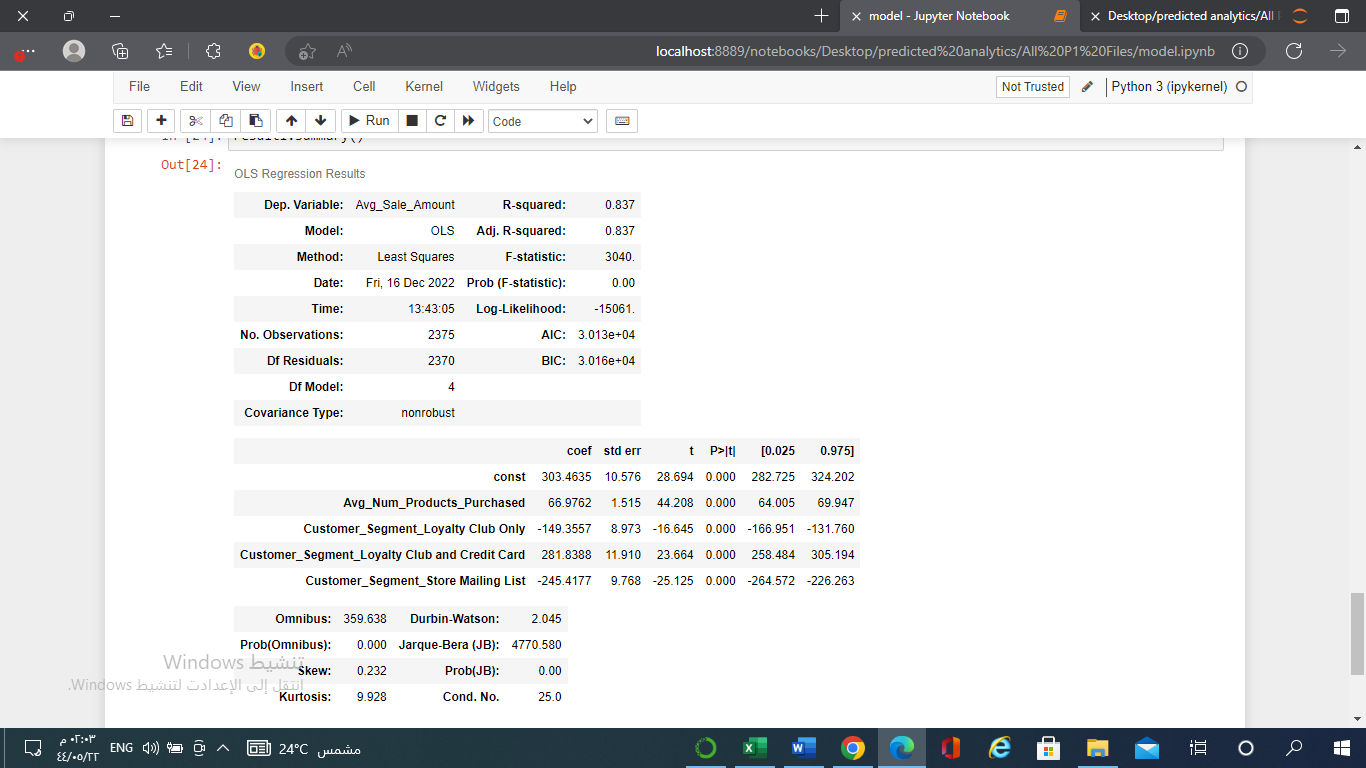
*At the minimum, answer these questions:*

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.

* The first variable I choose is *Avg\_Num\_Products\_Purchased* it is numeric variable so I did scatter plot to see if it has linear relationship with *Avg\_Sale\_Amount* and it indeed has linear relationship see the scatter plot



* The second variable I choose is *Customer\_Segment* and because it is categorical variable, I add it to the linear regression model and there was enhance in R-square and adjust R-square from 0.73 to 0.83 and the p-value was 0.00 which mean it is statistically significant and didn't occur by chance.



1. 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.

* At first, I did the model using one predictor variable which is *Avg\_Num\_Products\_Purchased the model achieve* 0.73 R-square and adjust R-square. *Avg\_Num\_Products\_Purchased* has p-value of 0.00 which mean it is statistically significant
* After that I added the second predictor variable which is *Customer\_Segment* and its good fit for the model because the R-square and adjust R-square increased from 0.73 to 0.83. *Customer\_Segment has p-value of 0.00 which mean it is statistically significant.*

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)

**Important: The regression equation should be in the form:**

**Avg\_Sale*\_per\_customer****=* 303.46 *+* 66.97 *\** **Avg\_Num\_Products\_Purchased +** 0*\** **Customer\_Segment Credit Card Only**- 149.35 *\** **Customer\_Segment\_Loyalty Club Only** *+* 281.83 *\** **Customer\_Segment\_Loyalty Club and Credit Card** - 245.41 *\**  **Customer\_Segment\_Store Mailing List**

## **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, I recommend the company to send the catalog to the 250 customers.

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

* I applied the linear regression equation to the 250 customers to get the average sales amount to each customer then I calculate the expected revenue by multiply average sales amount per customer by the probability that customer will buy our product (Score\_Yes). Finally, I calculate the profit by take the expected revenue and multiply it by the average gross margin 0.50 and subtract it by $6.5 which is the cost of distributing and printing per catalog

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

* The profit we expect from sending the catalog to the 250 customer is **$21986.15**

Before you Submit

Please check your answers against the requirements of the project dictated by the [rubric](https://review.udacity.com/#!/rubrics/186/view) here. Reviewers will use this rubric to grade your project.