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?

   To find out if sending a print catalog to 250 new customers will result in a profit that exceeds $10,000 after removing printing cost and gross margin.

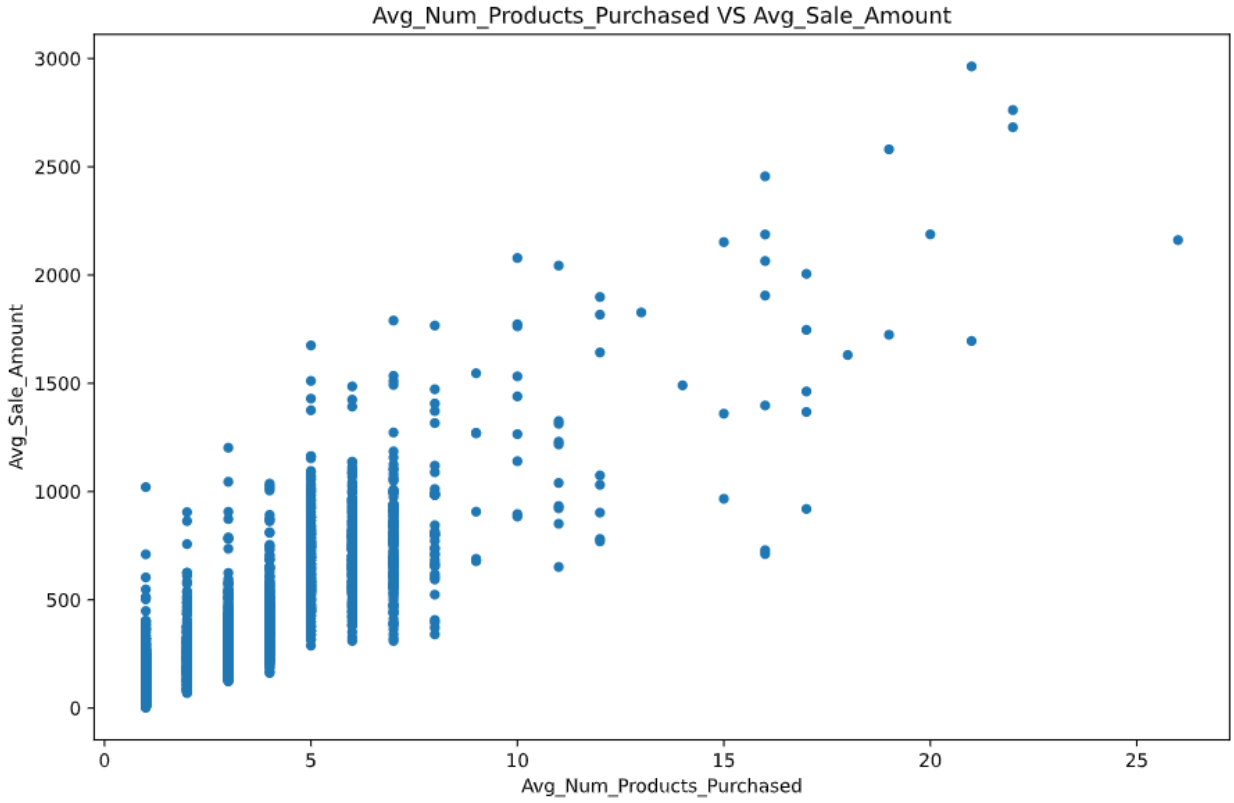
   The cost for printing is $6.50 per catalog and the average gross margin of 50%.

   Profit = total predicted sales * 0.5 - 6.5*250

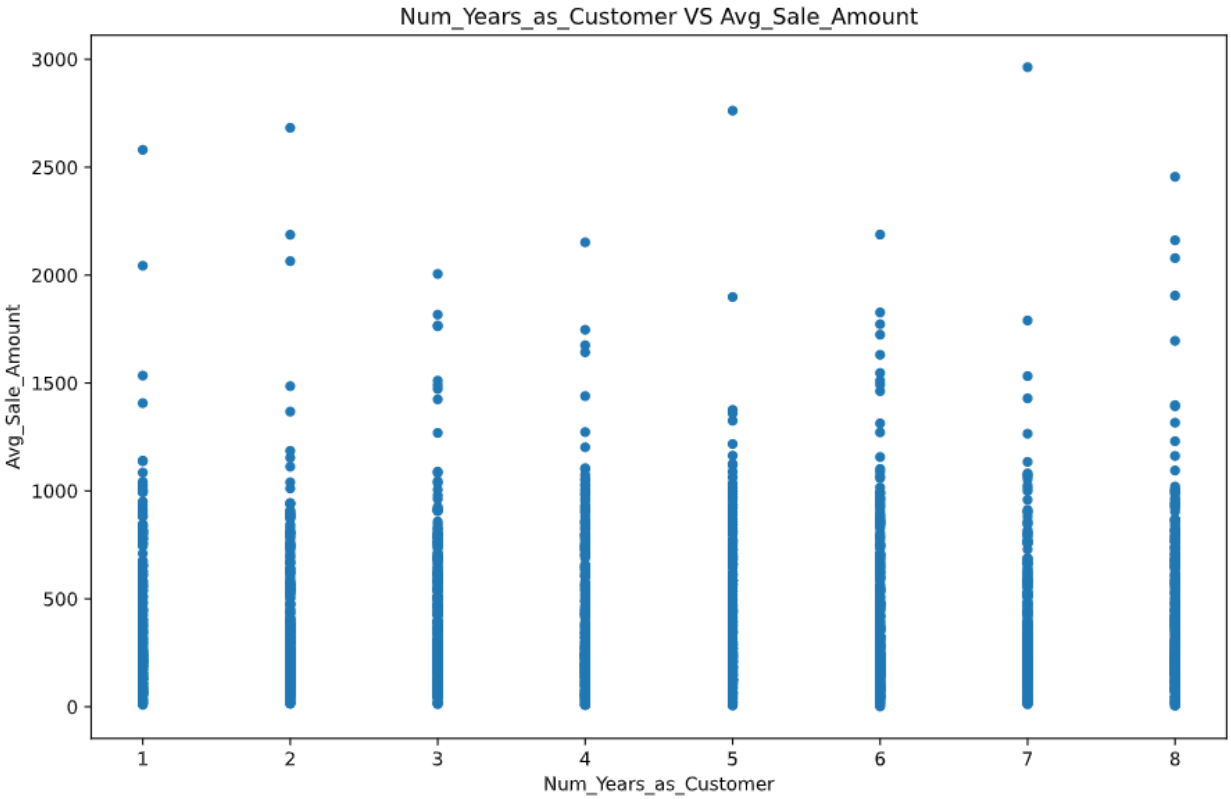
2. What data is needed to inform those decisions?
   The information Customer_Segment, Avg_Num_Products_Purchased and Num_Years_as_Customer can be used to find model for predicting Avg_Sale_Amount

Step 2: Analysis, Modeling, and Validation

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.

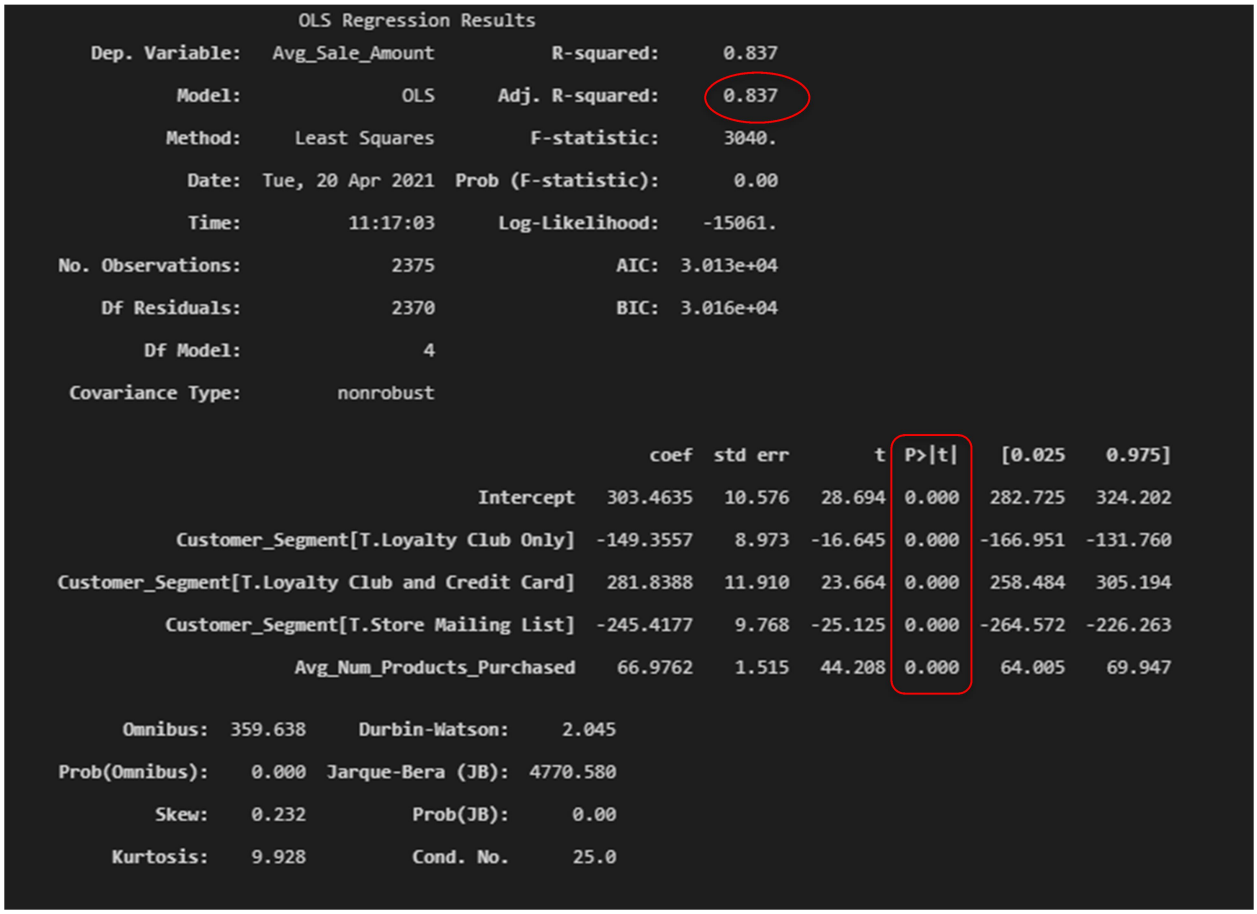


The variable Avg_Num_Products_Purchased show linear relashionship with target variable Avg_Sale_Amount.



The variable Num_Years_as_Customer show no linear relashionship with target variable Avg_Sale_Amount.

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.



OLS Regression Results

Dep. Variable:	Avg_Sale_Amount	R-squared:	0.837
Model:	OLS	Adj. R-squared:	0.837
Method:	Least Squares	F-statistic:	3040.
Date:	Tue, 20 Apr 2021	Prob (F-statistic):	0.00
Time:	11:17:03	Log-Likelihood:	-15061.
No. Observations:	2375	AIC:	3.013e+04
Df Residuals:	2370	BIC:	3.016e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	303.4635	10.576	28.694	0.000	282.725	324.202
Customer_Segment[T.Loyalty Club Only]	-149.3557	8.973	-16.645	0.000	-166.951	-131.760
Customer_Segment[T.Loyalty Club and Credit Card]	281.8388	11.910	23.664	0.000	258.484	305.194
Customer_Segment[T.Store Mailing List]	-245.4177	9.768	-25.125	0.000	-264.572	-226.263
Avg_Num_Products_Purchased	66.9762	1.515	44.208	0.000	64.005	69.947

Omnibus:	359.638	Durbin-Watson:	2.045
Prob(Omnibus):	0.000	Jarque-Bera (JB):	4770.580
Skew:	0.232	Prob(JB):	0.00
Kurtosis:	9.928	Cond. No.	25.0

The higher the adjusted R-squared value, the higher the explanatory power of the model. This value represents the amount of variation in the target variable explained by the variation in the predictor variables.

Any model with an adjusted R-square value above 0.70 is considered to be a strong model. Our present linear model has a value of 0.837; hence it is a good model.

The lower the p-value, the greater the statistical significance of the observed difference. The p-value is a measure of the probability that an observed difference could have occurred just by random chance.

Any model with a p-value less than 0.05 are considered to have statistical significance. Our model has a 0 p-value in both Customer_Segment and Avg_Num_Products_Purchased.

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)

Avg_Sale_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) + 0(if Customer_Segment: Credit Card Only)

Step 3: Presentation/Visualization

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

- ✓ Yes. The company should send the catalog to these 250 customers

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

- ✓ Use customer segment and average number of products purchased to predict average sale amount in mail list customer.
- ✓ Find possible amount customer spend by multiply predicted average sale and probability customer to buy (score_yes)
- ✓ Find total of all possible amount customer can spend
- ✓ To get profit, multiply total all possible amount customers can spend by 0.5 and minus total print cost (6.5*250)
- ✓ If profit is more than $10,000 then sending catalog is advised to company otherwise no.

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

- ✓ The expected profit is $21,987.44