Project: Predicting Catalog Demand

Step 1: Business and Data Understanding
Key Decisions:

1. What decisions needs to be made?
   How much profit the company can expect from sending a catalog to these customers?

2. What data is needed to inform those decisions?
   The customer segment and the avg_number_products_purchased.

Step 2: Analysis, Modeling, and Validation

- The target variable is the avg sale amount while keeping the predictor variable as customer segment and avg number of products purchased. I have selected the target and predictor variables because, we have to determine the product sale amount irrespective of the location, but we have a customer segment where the product purchase gets varies.

- Please find the regression in below graphs
  i) Avg_sale-amount Vs Avg_num_Product purchased, we can see a positive linear regression
  ii) Avg sale amount vs customer segment, we can see a positive linear regression.

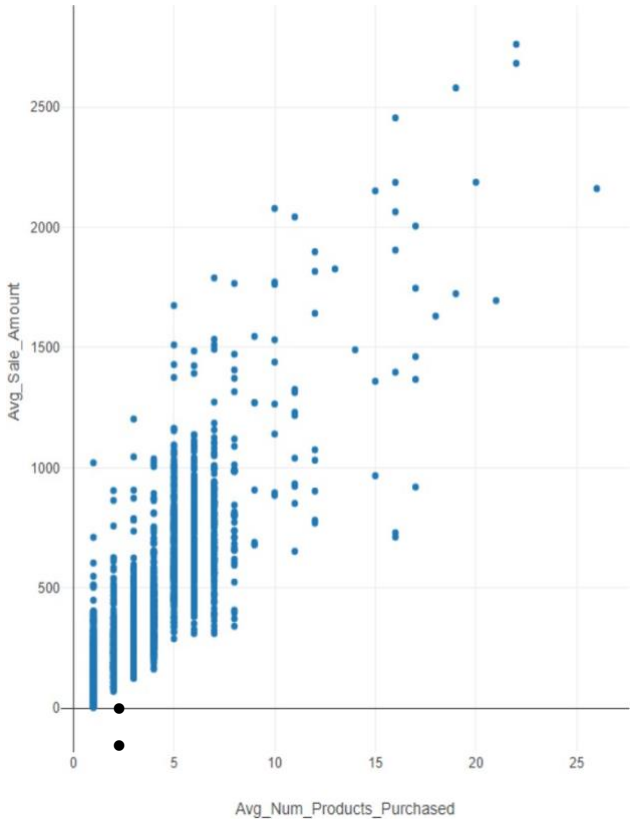
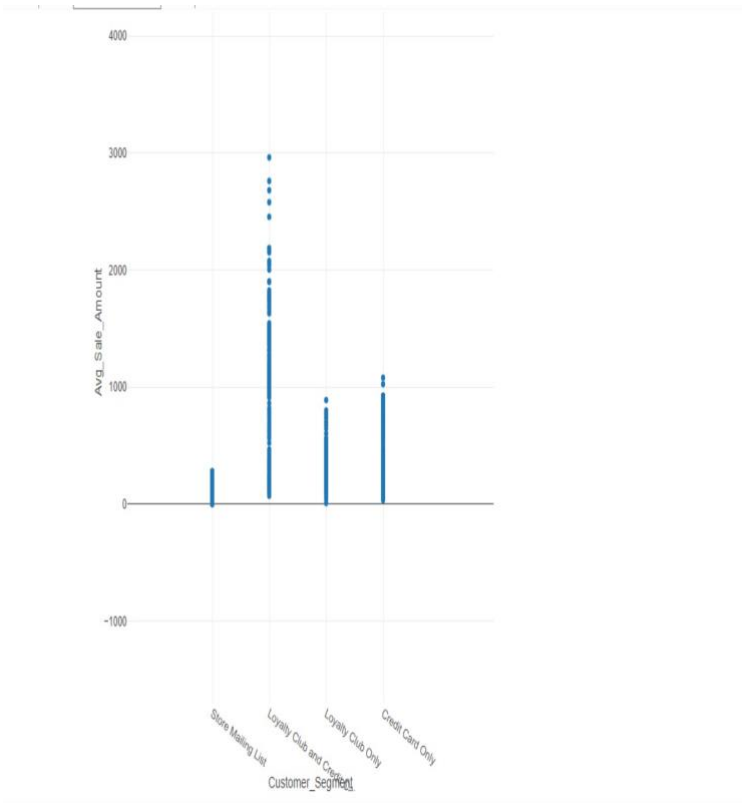


Fig (i)



Fig(ii)

- It has been asked whether the mailing list of 250 customers will be helpful in earning profit beyond 10000$ irrespective of location.
- I have selected the customer segment and avg number of products purchased field to analyze keeping the avg sale amount field as the target variable.
- It is because, while selecting the other parameters like store id, city gives the p value above 0.05 and it is somewhat irrelevant since there is no false data present in the data set.
- P value and r- squared values for the model are 2.2e-16 and 0.8366 respectively.

Important: The regression equation should be in the form:

*Y = Intercept + b1 * Variable_1 + b2 * Variable_2 + b3 * Variable_3……*
*Y= 303.46 + 66.98 *** Avg_Num_Products_Purchased - 149.36 * Loyalty Club Only + 281.84 * Loyalty Club and Credit Card - 245.42 * Store Mailing List + 0 * Credit Card Only

Note that we **must** include the 0 coefficient for the type Credit Card Only.

Report for Linear Model Customer_Relation

Basic Summary

Call:
lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = the.data)

Residuals:

Min	1Q	Median	3Q	Max
-663.8	-67.3	-1.9	70.7	971.7

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	303.46	10.576	28.69	< 2.2e-16 ***
Customer_SegmentLoyalty Club Only	-149.36	8.973	-16.65	< 2.2e-16 ***
Customer_SegmentLoyalty Club and Credit Card	281.84	11.910	23.66	< 2.2e-16 ***
Customer_SegmentStore Mailing List	-245.42	9.768	-25.13	< 2.2e-16 ***
Avg_Num_Products_Purchased	66.98	1.515	44.21	< 2.2e-16 ***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 137.48 on 2370 degrees of freedom
Multiple R-squared: 0.8369, Adjusted R-Squared: 0.8366
F-statistic: 3040 on 4 and 2370 degrees of freedom (DF), p-value < 2.2e-16

Type II ANOVA Analysis

Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)
Customer_Segment	28715078.96	3	506.4	< 2.2e-16 ***
Avg_Num_Products_Purchased	36939582.5	1	1954.31	< 2.2e-16 ***
Residuals	44796869.07	2370		

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step 3: Presentation/Visualization

1. The recommendation. Should the company send the catalog to these 250 customers?
   Company can send the catalog to these 250 customers because it will be generating a revenue of $47224.87 with a margin of $23612.43 and if we substract the catalog price then the profit will be (23612-(250*6.5))= $21987.43

2. Reason for recommendation?
   Since the company wants to send the catalog only if the profit exceeds $ 10000 and it is satisfying the condition.

3. Expected profit from the new catalog (assuming the catalog is sent to these 250 customers)
   The expected profit is of $23612.43 and if we substract the catalog price then the profit will be (23612-(250*6.5))= $ 21987.43