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?

Answer:

The business problem is about predicting the expected profit the company will made when catalog are send to new 250 customers. Predictive analytics is needed or ensured so the business problems could be solved. As I realised there were a lot of data from the business problem that needed to be analysis, so I decided to apply data-rich predictive analysis that could help me predict the outcome of the data. Since the outcome involves numbers, I applied numeric predictive analysis and immediately I realised the outcome is continuous numeric variables involving range of numbers. After I got the prediction right, I applied the linear regression to build the model so it can facilitates the prediction of sales for each customer from their mailing_list. In conclusion, the predicting sales was used to produce the expected profit so it can be decided if the company must send the catalog to the new 250 customers or not.

2. What data is needed to inform those decisions?

Answer:

The data on customer_ID and number (#) of years as a customer confirming whether they have bought or received catalog from the previous and current year(s). Again the transactional (product purchases and sales) data of customers from the past. The revenue = price-margin – cost helps in calculating the expect profit that a customer spent on purchasing from the company catalog.

Step 2: Analysis, Modelling, 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)

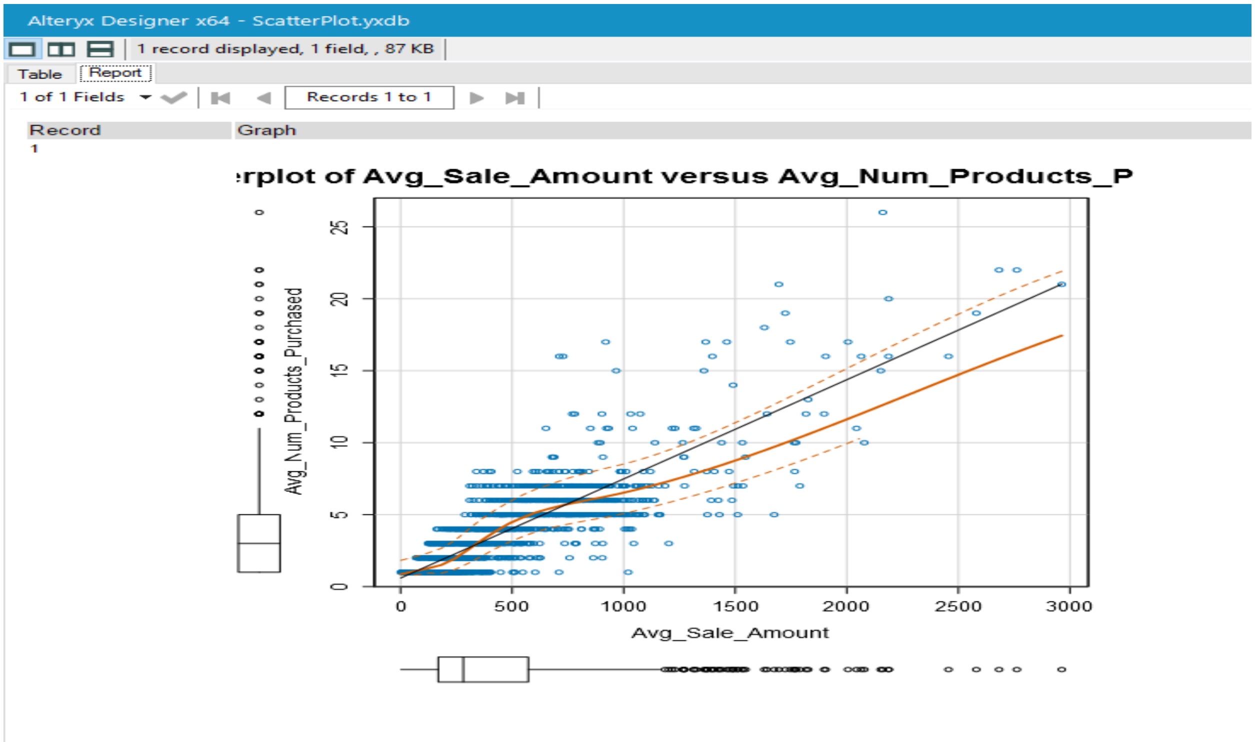
Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

1. How and why did you select the predictor variables (see supplementary text) in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer to this lesson to help you explore your data and use scatterplots to search for linear relationships. You must include scatterplots in your answer.

Answer:

In order to build up linear regression model, the file p1-customers.xlsx was selected as the data file and the predictor variables entail the customer segment and avg_num_products purchased. It must be noted that purchases of catalog made in the past are not important to the new customers and the others (categorical variables) have no influences on new purchases. The avg_sales_amount is the target variable which help in scoring the p1-mailinglist.xlsx and the probability of buying catalog is the score_yes which is on the x axis and score on y axis.


Alteryx Designer x64 - ScatterPlot.yxdb

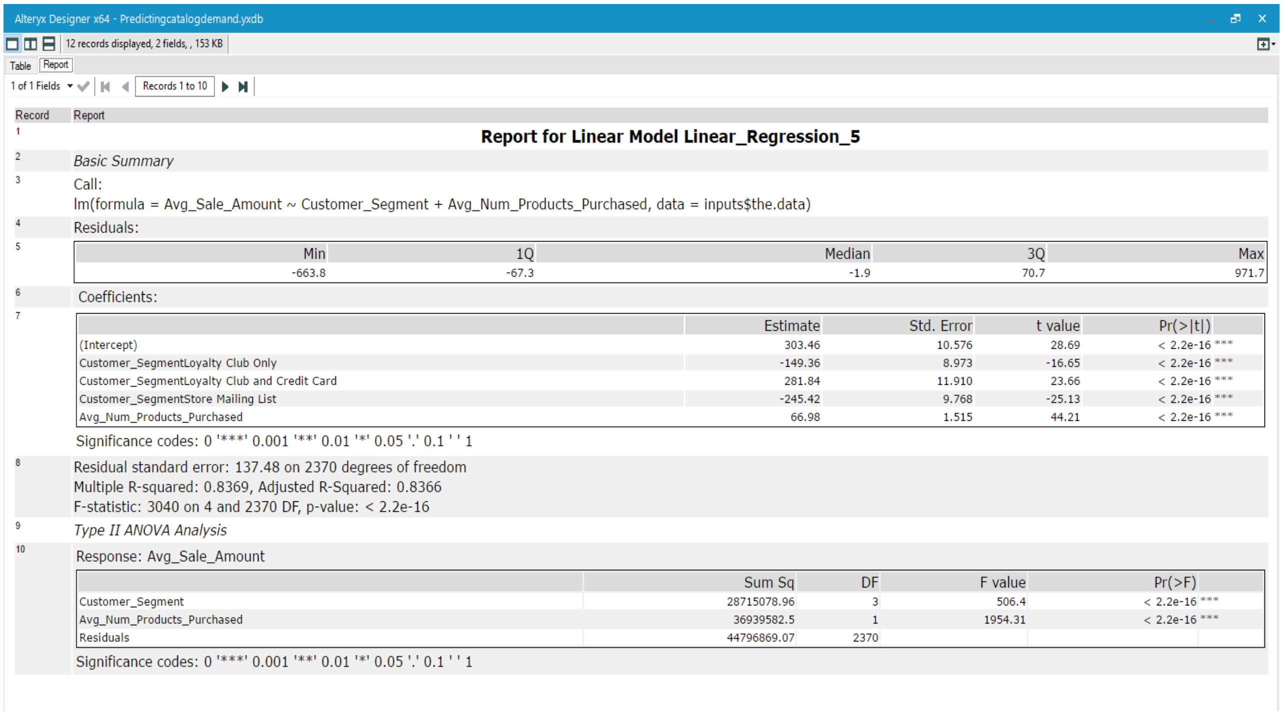
1 record displayed, 1 field, , 87 KB

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.

Answer:

Since the multiple R-squared and the adjusted R-squared are 0.8369 and 0.8366 respectively and also the p-value is < 2.2e-16. I can confidently confirm my linear model is a good model since the p-value has a lot of stars (***) meaning statistical significance and the R-squared ranges from 0 to 1 that represents the amount of variation in the target variable explained by the variation in the predictor variables. The R-squared value of 0.8 is approximately 1 which represents a good fit, making it a strong model too.

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)
Answer:
Y= 303.46 – 149.36*cust_segmentLoyalty club + 281.84*cust_segmentLoyalty club & credit card – 245.42*cust_segmentStore Mailing List + 66.98*Avg_num_products_purchased + 0*credit_card.



Alteryx Designer x64 - Predictingcatalogdemand.yxdb
12 records displayed, 2 fields, , 153 KB
Table
Report
1 of 1 Fields
Records 1 to 10

Record	Report
1	**Report for Linear Model Linear_Regression_5**
2	*Basic Summary*
3	Call: lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = inputs$the.data)
4	Residuals:
5	

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

6 Coefficients:

7

	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

8 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 DF, p-value: < 2.2e-16

9 *Type II ANOVA Analysis*

10 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

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?

Answer:

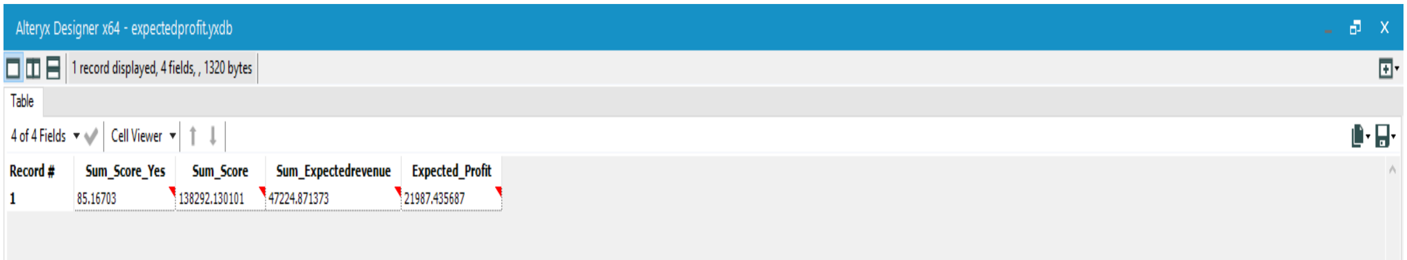
I think it would be reasonable for the company to send the catalog to the 250 customers since the expected profit ($21987.435687) is more than the $10,000.

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

Answer:

- I calculated for the expected revenue which is score*score_yes. Then the expected revenue was attained.
- I calculated for the expected profit using the total of expected revenue* average gross margin of 0.5 subtracted by the cost of printing and distributing per catalog * the 250 customers. Simply, the expected profit = (total sum of expected revenue * 0.5) – (6.5*250).

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



1 record displayed, 4 fields, , 1320 bytes

Record #	Sum_Score_Yes	Sum_Score	Sum_Expectedrevenue	Expected_Profit
1	85.16703	138292.130101	47224.871373	21987.435687