Predictive Analytics for Business

Project#1 Predicting Diamond Prices

Name: Marwan Saeed Alsharabbi

Date: 29-11-2019

2019

INTRODUCTION

.

The Business Problem

You recently started working for a company that manufactures and sells high-end home goods. Last year the company sent out its first print catalog, and is preparing to send out this year's catalog in the coming months. The company has 250 new customers from their mailing list that they want to send the catalog to.

Your manager has been asked to determine how much profit the company can expect from sending a catalog to these customers. You, the business analyst, are assigned to help your manager run the numbers. While fairly knowledgeable about data analysis, your manager is not very familiar with predictive models.

You've been asked to predict the expected profit from these 250 new customers. Management does not want to send the catalog out to these new customers unless the expected profit contribution exceeds $10,000.

Details

The costs of printing and distributing is $6.50 per catalog.
The average gross margin (price - cost) on all products sold through the catalog is 50%.
Make sure to multiply your revenue by the gross margin first before you subtract out the $6.50 cost when calculating your profit.
Write a short report with your recommendations outlining your reasons why the company should go with your recommendations to your manager.

Steps to Success

Step 1: Business and Data Understanding

Your project should include:

A description of the key business decisions that need to be made.

Note: Clean data is provided for this project, so you can skip the data preparation step of the Problem Solving Framework.

Step 2: Analysis, Modeling, and Validation

Build a linear regression model, then use it to predict sales for the 250 customers. We encourage you to use Alteryx to build the best linear model with your data.

Note: For students using software other than Alteryx, if you decide to use Customer Segment as one of your predictor variables, please set the base case to Credit Card Only.
However, feel free to use any tool you'd like. You should create your linear regression model and come up with a linear regression equation.

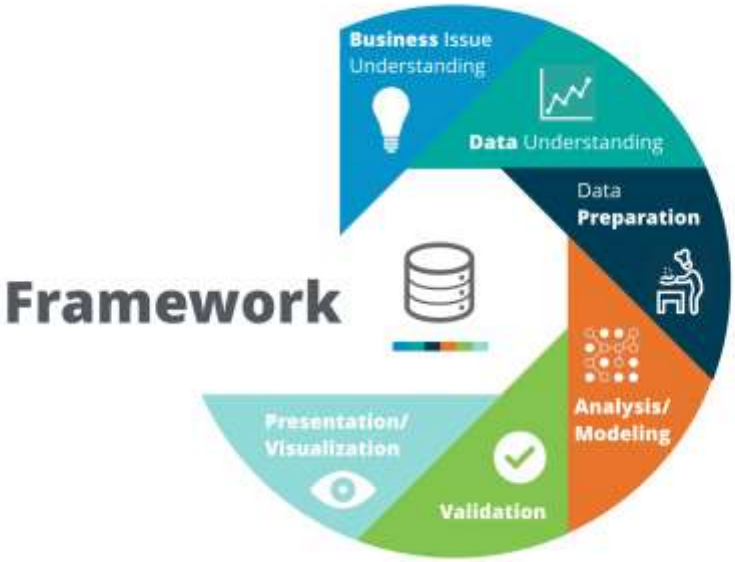
Once you have your linear regression equation, you should use your linear regression equation to predict sales for the individual people in your mailing list.

Step 3: Writeup

Once you have your predicted or expected profit, write a brief report with your recommendation to whether the company should send the catalog or not.
Hint: We want to calculate the expected revenue from these 250 people in order to get expected profit. This means we need to multiply the probability that a person will buy our catalog as well. For example, if a customer were to buy from us, we predict this customer will buy $450 worth of products. At a 30% chance that this person will actually buy from us, we can expect revenue to be $450 x 30% = $135.

What is the Cross Industry Standard Process for Data Mining (CRISP-DM)

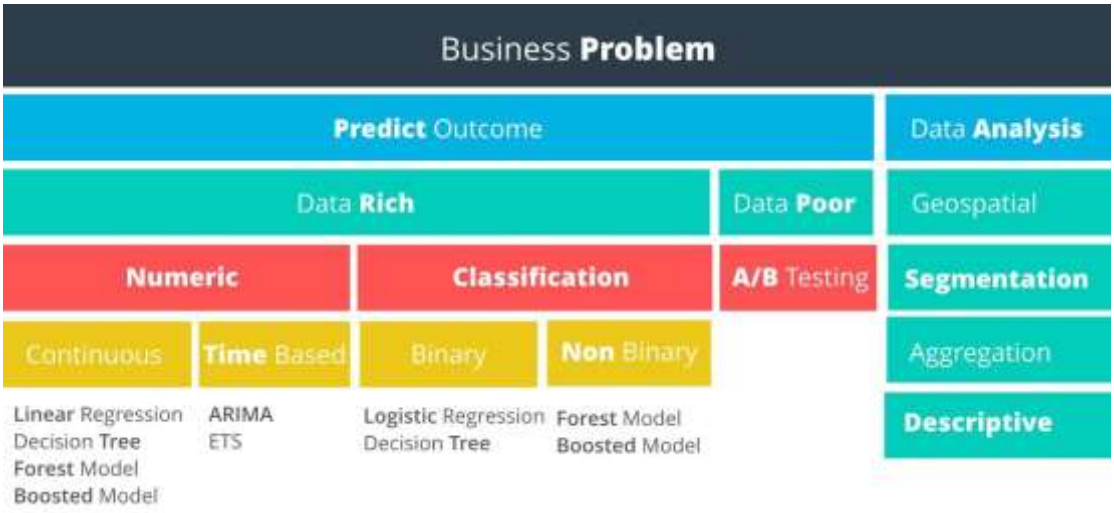
"A data mining process model that describes commonly used approaches that data mining experts use to tackle problems... it was the leading methodology used by industry data miners.



What is the Methodology Map?

The Analytical Problem Solving framework helps systematically work through a business problem; however, it doesn't help us understand which methodology to use.

The Methodology Map will help us as we decide which methodology to use to solve a business problem, and is meant to be used as we work our way through the Problem Solving Framework



What is the linear regression?

Linear Regression is a statistical method used to predict numeric outcomes by analyzing the outcome's relationship with one or more predictor variables.

Simple Regression

A regression with only one explanatory or X variable

Multiple Regression

A regression with two or more explanatory or X variables

Linear Regression for Business Statistics

Overview of Regression

- *Modeling Developing a regression model.*

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$$

Y	**X1,X2,X3**	**B1,B2,B3**
Target Variable Variable of interest Y variable Dependent variable Response variable Regressed variable L.H.S. variable	Predictor Variable Explanatory variables X variables Independent variables Covariate variables Repressor variables **R.H.S" variables**	**Coefficients Parameters** **B0= Y-intercept**

- Estimation *Using software to estimate the model.*
- Inference *Interpreting the estimated regression model .*
- Prediction *Making predictions about the variable of interest*

I will not give more details on how to create the model because it already exists will carry out the required steps for the project

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?
   Sending this year's catalog out to 250 new customers from the mailing list
   The business decision here is to decide whether to go with a new catalogue launch for their high-end home goods product range. There is a cost to producing these Catalogues and an expected profit margin that needs to be obtained.
   The decision to need to be made is that we should be sending the catalog to new 250 customers if the profit will exceed over 10.000$.

2. What data is needed to inform those decisions?
   First, we need previous customer data so we could decide on the target and predictor variable. In this case we need to calculate the expected revenue from new 250 customers for getting the expected profit. So, we will multiply the predicted sales amount (Score) by probability to buy (Score_Yes) for getting the predicted sales. As provided, we will multiply the predicted sales by 50% then subtract out the $6.50 cost for getting the predicted profit
   The data needed will be building the model linear regression
   • Avg_Sale_Amount
   • Avg_Num_Products_Purchased
   • Years_As_Customer
   • Store_Number
   • Customer_Segment

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)

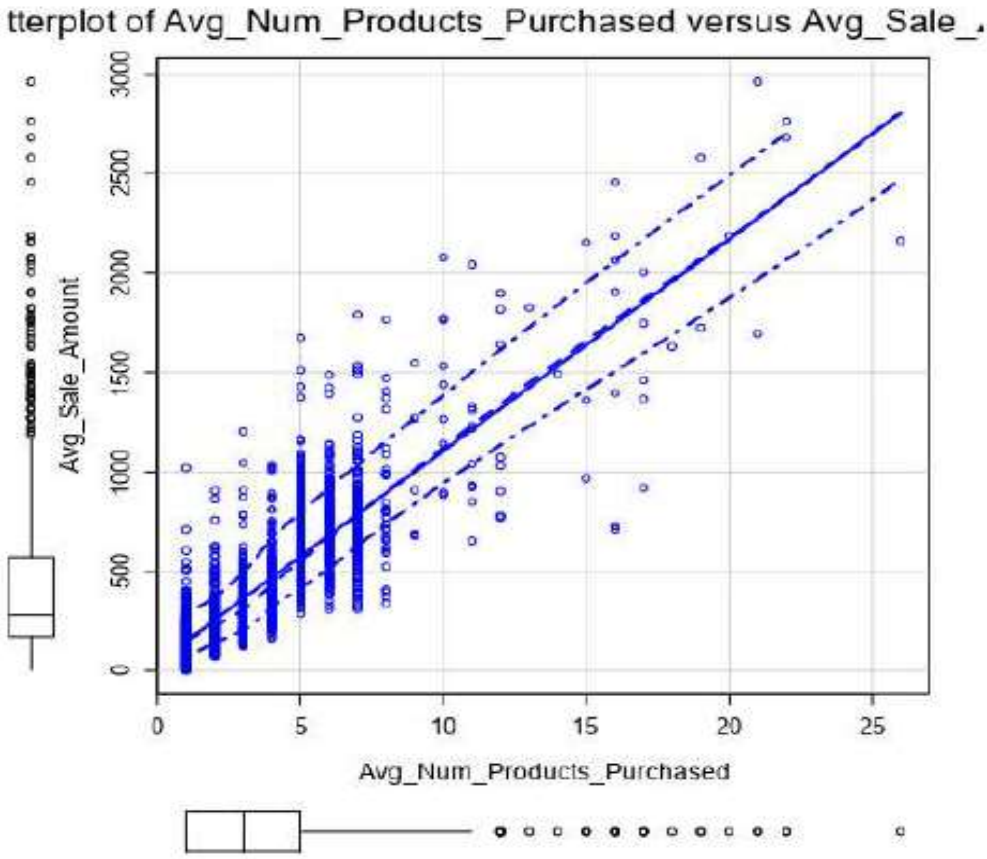
Important: Use the p1-customers.xlsx to train your linear model.

At the minimum, answer these questions:

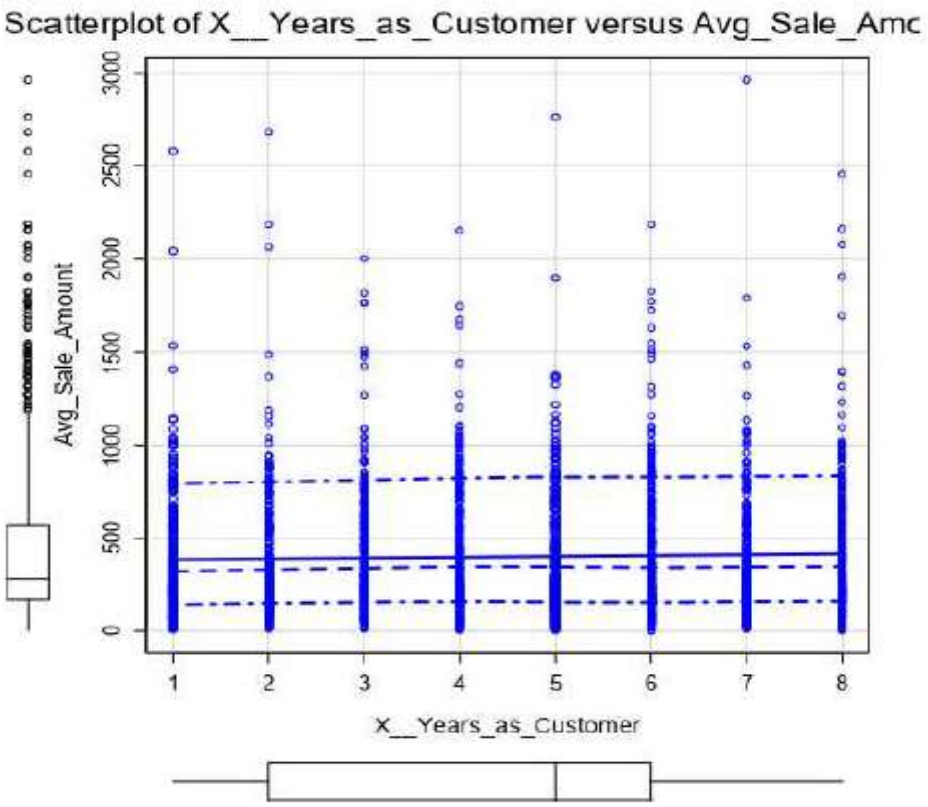
1. How and why did you select the predictor variables in your model? You must explain how your continuous predictor variables you've chosen have a linear relationship with the target variable. Please refer back to the "Multiple Linear Regression with Excel" lesson to help you explore your data and use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

   I chose Customer_Segment as the non-numeric predictor variable. Plus, I initially chose Avg_Num_Products_Purchased and Avg_Sale_Amount see (Plot#1) noted the strong correlation and selected Avg_Sale_Amount and Numeber_Years_as_Customer see (Plot#2) and too Customer_Segment, Avg_Sale_Amount see (Plot#3)

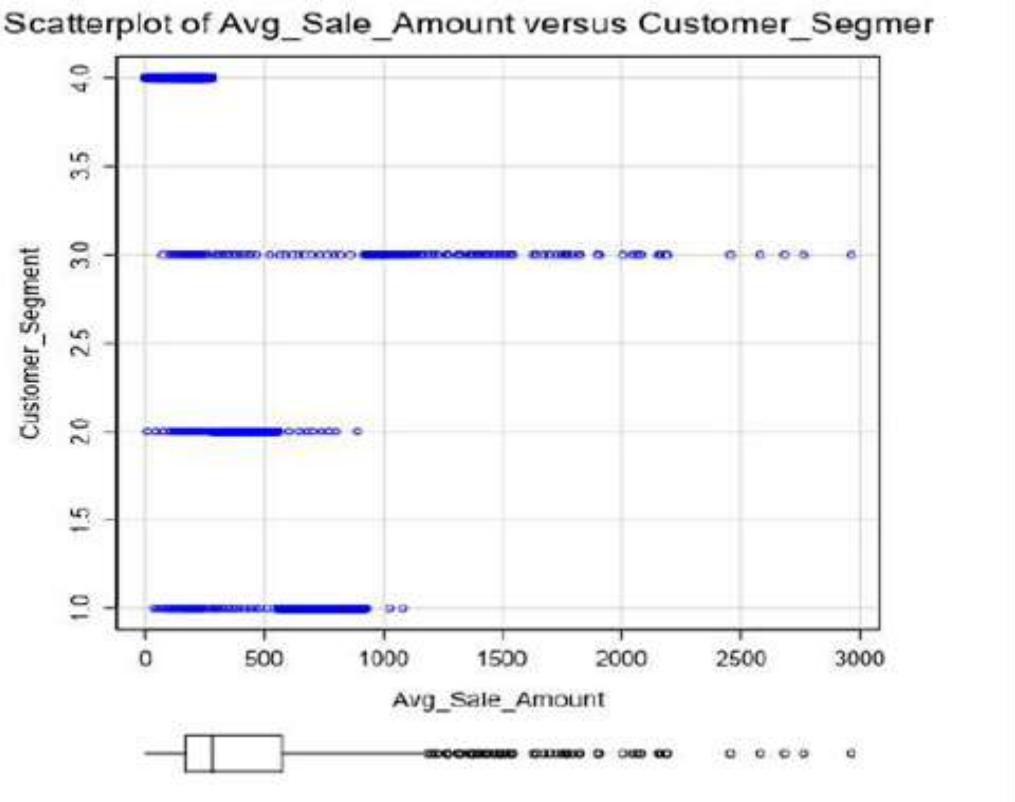
   Because both have continuous numeric values and are likely to help predict sales look to the scatterplots below



Plot#1 Scatterplot of average sale amount versus average number of products purchased. There is a positive l linear relationship is strong (positive slope).



Plot#2 Scatterplot of average sale amount versus years as customer. There is almost no relationship between the variables, as it is indicated by a slope close to zero.

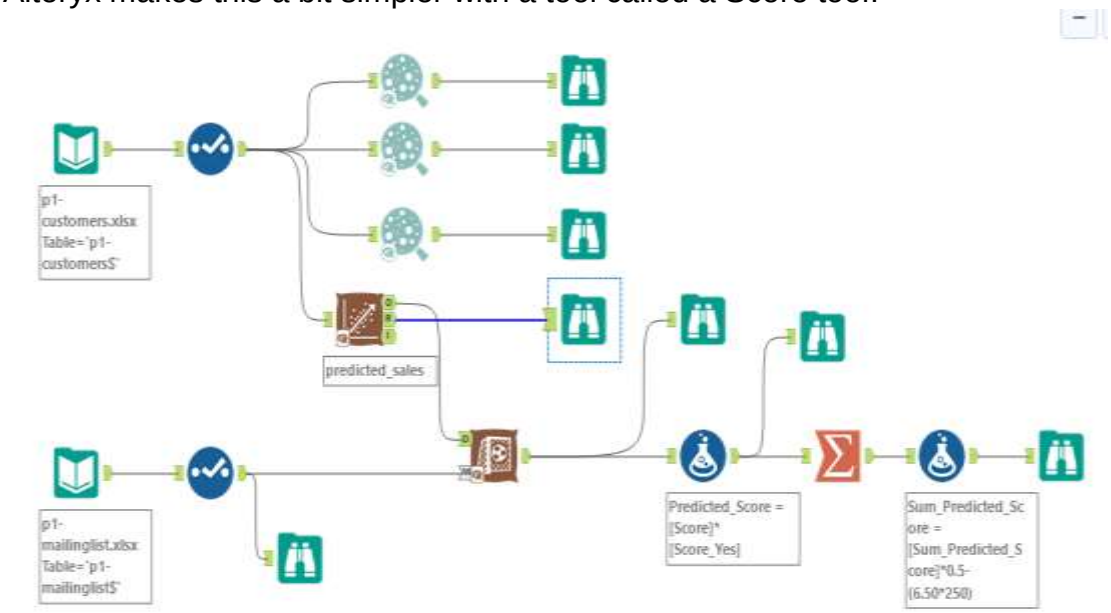


Plot#3

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.

Now that the data needed will be building the model linear regression, it's time to apply the Company has 250 new customers from their mailing list that they want to send the catalog to.

Prediction and recommendation. For project zero you did this manually in a spreadsheet by using the linear regression equation and plugging in values. Alteryx makes this a bit simpler with a tool called a Score tool.



Because the good model Show the results in detail as follows

Report for Linear Model predicted_sales

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

1. What price do you recommend the jewelry company to bid? Please explain how you arrived at that number.

Based on the predictive model's findings, and taking into account the jewelry company Generally purchases diamonds from distributors at 70% of the retail price, I used the formula from the linear regression model (based on the characteristics and prices of the previous diamond sales), applied it to the 3,000 diamonds and then I added up all those predicted prices. Finally, I multiplied that amount ($11733522.76) by 0.70 to get the final predicted bid of $ 8213465.932 bid for the new set of 3,000 diamonds is ~ $ 8213465.932

I wish success to all.

Marwan Saeed Alsharabbi