Project 1: Predicting Catalog Demand

Step 1: Business and Data Understanding

Key Decisions:

1. What decisions needs to be made?

The decision that needs to be made is if is it worth sending a catalog to 250 new customers? only if these will generate a total profit superior at $10,000.

2. What data is needed to inform those decisions?

1st. We need to look at the linearity between all the available predictor variables and the target variable (average sales amount) to choose this or those that show a good correlation bearing in mind sadistic analysis.

2nd. After finding the right predictor variable or variables, we need to build the predictor model and get the equation to predict the average sale amount by the customer.

3rd. After calculated the predicted average sales amount using the equation of the predictive model for every one of the 250 customers, we need to calculate the predictive profit by customer bearing in mind that:

- **The probability of buying after the customer had received the catalog (data gave us in the field Scores_Yes)**
- **Considering that the gross margin by-product is of 50% of the price of sales for all the products sold by catalog**
- **And that the catalog's cost of printing and distribution is $6.50**

Somehow the predicted profit by a new customer is:

$$(Predictive\_Avg\_Sales\_Amount * Score\_Yes * gross\_margin\ ) - cost\_catalog\_printing\_and\_distribution$$

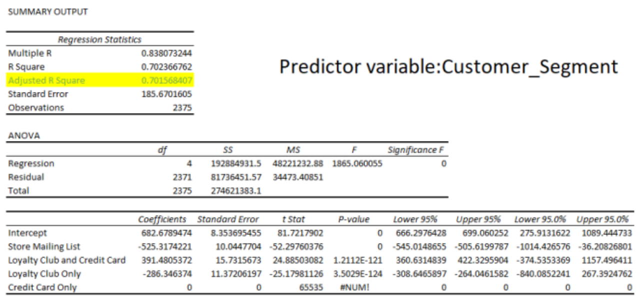
4th. Finally, to get the total projected profit we need to sum the predicted profit by each one new customer that has received the catalog and to take our decision to send or not the catalog, regard the manager's advice

Step 2: Analysis, Modeling, and Validation

2. 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.

1st. I had determinate what model I need to run based on the data that I had and chose the right model to predict the average sales amount. I had enough data; my output must be a number and for the type of data the suggested model is the "Linear Regression"

2nd. I generated a scatter plot to visualize data correlation between predictor variables and the target variable. I used Microsoft regression analysis tool for the different predictor variables, I chose the predictor variable with an adjusted r^2 equal to or greater than 0.7



SUMMARY OUTPUT

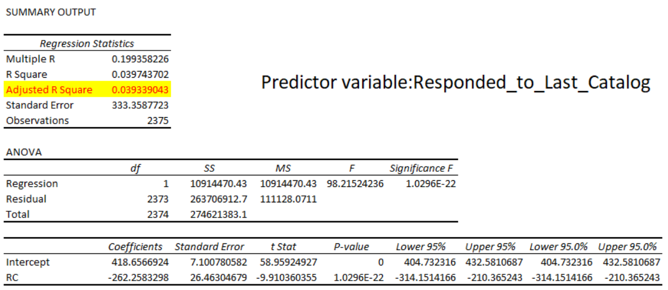
Regression Statistics	
Multiple R	0.838073244
R Square	0.702366762
Adjusted R Square	0.701568407
Standard Error	185.6701605
Observations	2375

Predictor variable:Customer_Segment

ANOVA

	df	SS	MS	F	Significance F
Regression	4	192884931.5	48221232.88	1865.060055	0
Residual	2371	81736451.57	34473.40851		
Total	2375	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	682.6789474	8.353695455	81.7217902	0	666.2976428	699.060252	275.9131622	1089.444733
Store Mailing List	-525.3174221	10.0447704	-52.29760376	0	-545.0148655	-505.6199787	-1014.426576	-36.20826801
Loyalty Club and Credit Card	391.4805372	15.7315673	24.88503082	1.2112E-121	360.6314839	422.3295904	-374.5353369	1157.496411
Loyalty Club Only	-286.346374	11.37206197	-25.17981126	3.5029E-124	-308.6465897	-264.0461582	-840.0852241	267.3924762
Credit Card Only	0	0	65535	#NUM!	0	0	0	0



SUMMARY OUTPUT

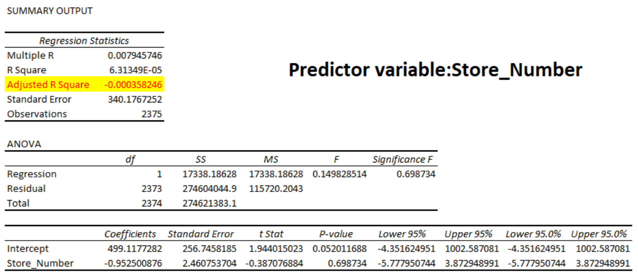
Regression Statistics	
Multiple R	0.199358226
R Square	0.039743702
Adjusted R Square	0.039339043
Standard Error	333.3587723
Observations	2375

Predictor variable:Responded_to_Last_Catalog

ANOVA

	df	SS	MS	F	Significance F
Regression	1	10914470.43	10914470.43	98.21524236	1.0296E-22
Residual	2373	263706912.7	111128.0711		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	418.6566924	7.100780582	58.95924927	0	404.732316	432.5810687	404.732316	432.5810687
RC	-262.2583298	26.46304679	-9.910360355	1.0296E-22	-314.1514166	-210.365243	-314.1514166	-210.365243



SUMMARY OUTPUT

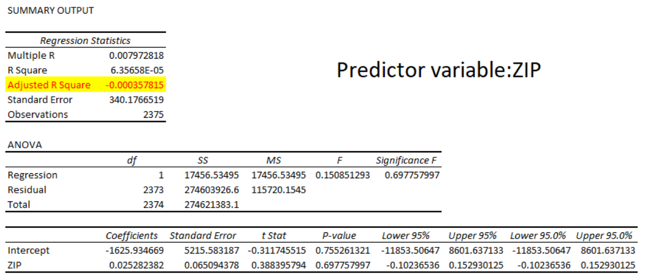
Regression Statistics	
Multiple R	0.007945746
R Square	6.31349E-05
Adjusted R Square	-0.000358246
Standard Error	340.1767252
Observations	2375

Predictor variable:Store_Number

ANOVA

	df	SS	MS	F	Significance F
Regression	1	17338.18628	17338.18628	0.149828514	0.698734
Residual	2373	274604044.9	115720.2043		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	499.1177282	256.7458185	1.944015023	0.052011688	-4.351624951	1002.587081	-4.351624951	1002.587081
Store_Number	-0.952500876	2.460753704	-0.387076884	0.698734	-5.777950744	3.872948991	-5.777950744	3.872948991



SUMMARY OUTPUT

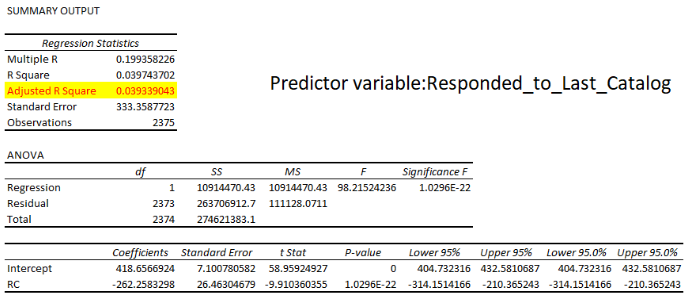
Regression Statistics	
Multiple R	0.007972818
R Square	6.35658E-05
Adjusted R Square	-0.000357815
Standard Error	340.1766519
Observations	2375

Predictor variable:ZIP

ANOVA

	df	SS	MS	F	Significance F
Regression	1	17456.53495	17456.53495	0.150851293	0.697757997
Residual	2373	274603926.6	115720.1545		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1625.934669	5215.583187	-0.311745515	0.755261321	-11853.50647	8601.637133	-11853.50647	8601.637133
ZIP	0.025282382	0.065094378	0.388395794	0.697757997	-0.10236536	0.152930125	-0.10236536	0.152930125



SUMMARY OUTPUT

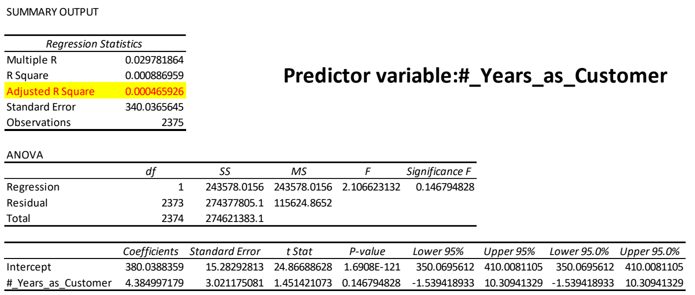
Regression Statistics	
Multiple R	0.199358226
R Square	0.039743702
Adjusted R Square	0.039339043
Standard Error	333.3587723
Observations	2375

Predictor variable:Responded_to_Last_Catalog

ANOVA

	df	SS	MS	F	Significance F
Regression	1	10914470.43	10914470.43	98.21524236	1.0296E-22
Residual	2373	263706912.7	111128.0711		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	418.6566924	7.100780582	58.95924927	0	404.732316	432.5810687	404.732316	432.5810687
RC	-262.2583298	26.46304679	-9.910360355	1.0296E-22	-314.1514166	-210.365243	-314.1514166	-210.365243



SUMMARY OUTPUT

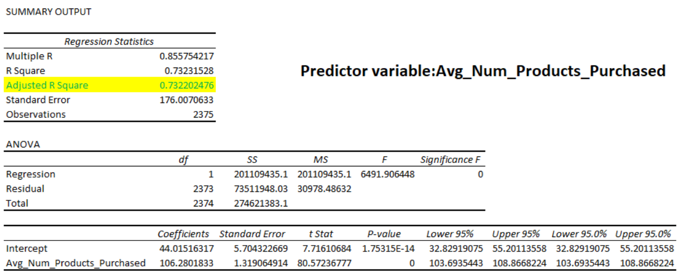
Regression Statistics	
Multiple R	0.029781864
R Square	0.000886959
Adjusted R Square	0.000465926
Standard Error	340.0365645
Observations	2375

Predictor variable:#_Years_as_Customer

ANOVA

	df	SS	MS	F	Significance F
Regression	1	243578.0156	243578.0156	2.106623132	0.146794828
Residual	2373	274377805.1	115624.8652		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	380.0388359	15.28292813	24.86688628	1.6908E-121	350.0695612	410.0081105	350.0695612	410.0081105
#_Years_as_Customer	4.384997179	3.021175081	1.451421073	0.146794828	-1.539418933	10.30941329	-1.539418933	10.30941329



SUMMARY OUTPUT

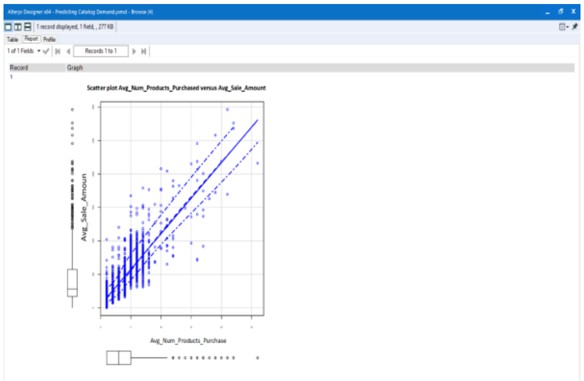
Regression Statistics	
Multiple R	0.855754217
R Square	0.73231528
Adjusted R Square	0.732202476
Standard Error	176.0070633
Observations	2375

Predictor variable:Avg_Num_Products_Purchased

ANOVA

	df	SS	MS	F	Significance F
Regression	1	201109435.1	201109435.1	6491.906448	0
Residual	2373	73511948.03	30978.48632		
Total	2374	274621383.1			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	44.01516317	5.704322669	7.71610684	1.75315E-14	32.82919075	55.20113558	32.82919075	55.20113558
Avg_Num_Products_Purchased	106.2801833	1.319064914	80.57236777	0	103.6935443	108.8668224	103.6935443	108.8668224



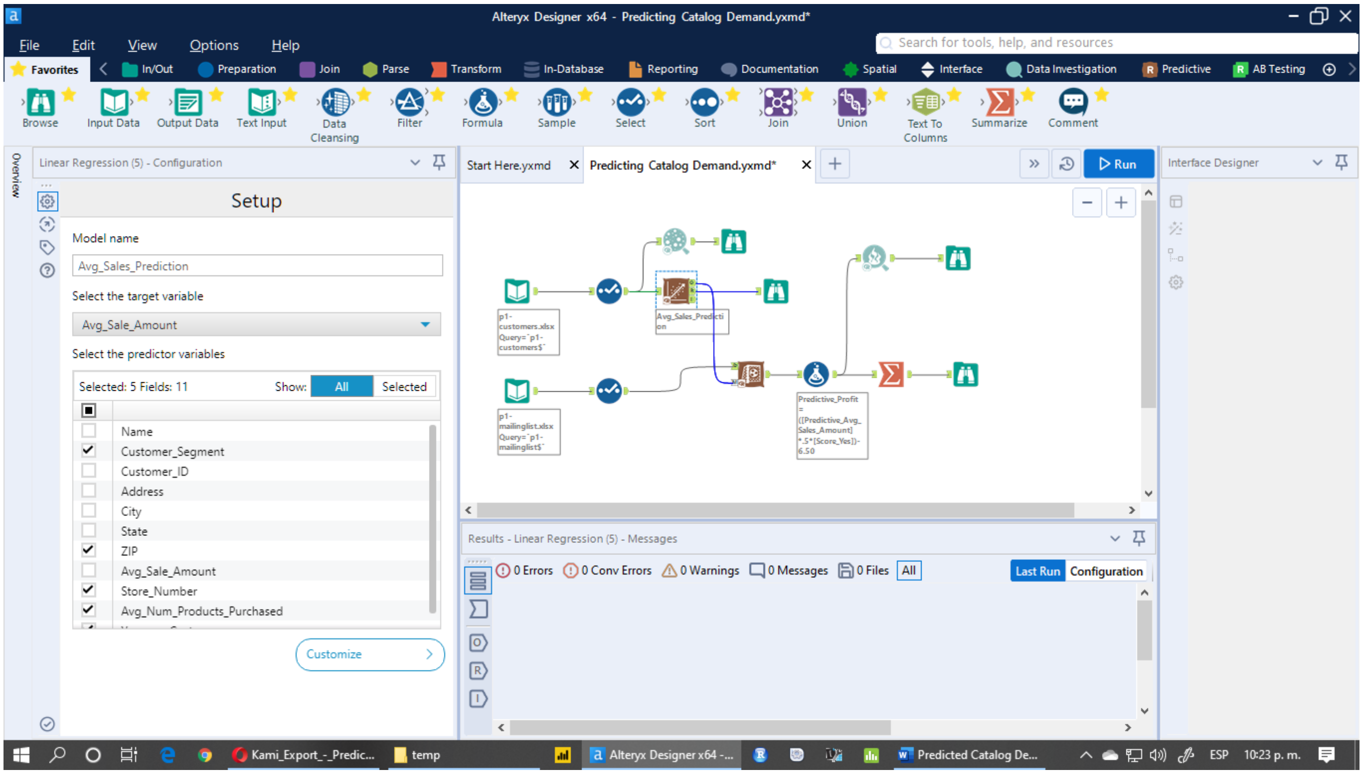
Alteryx Designer x64 - Predicting Catalog Demand.yxmd - Browse (4)

1 record displayed, 1 field, , 277 KB

Table
Report
Profile
1 of 1 Fields
Records 1 to 1
Record
Graph
1
Scatter plot Avg_Num_Products_Purchased versus Avg_Sale_Amount
Avg_Sale_Amoun
Avg_Num_Products_Purchase

3. 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.

1st. I ran the linear regression in Alteryx to found which are the predictor variables (Customer_Segment, ZIP, Store_Number, Avg_Num_Products_Purchased, #_Year_as_Customer, and Respond_to_las_Catalog) that model or predict in a significant way the target variable (Avg_Sales_Amount), founding that the Customer_Segment and Avg_Num_Products_Purchased are the predictor variables that show the best correlation



Linear Regression (5) - Configuration

Setup

Model name

Avg_Sales_Prediction

Select the target variable

Avg_Sale_Amount

Select the predictor variables

Selected: 5 Fields: 11 Show: All Selected

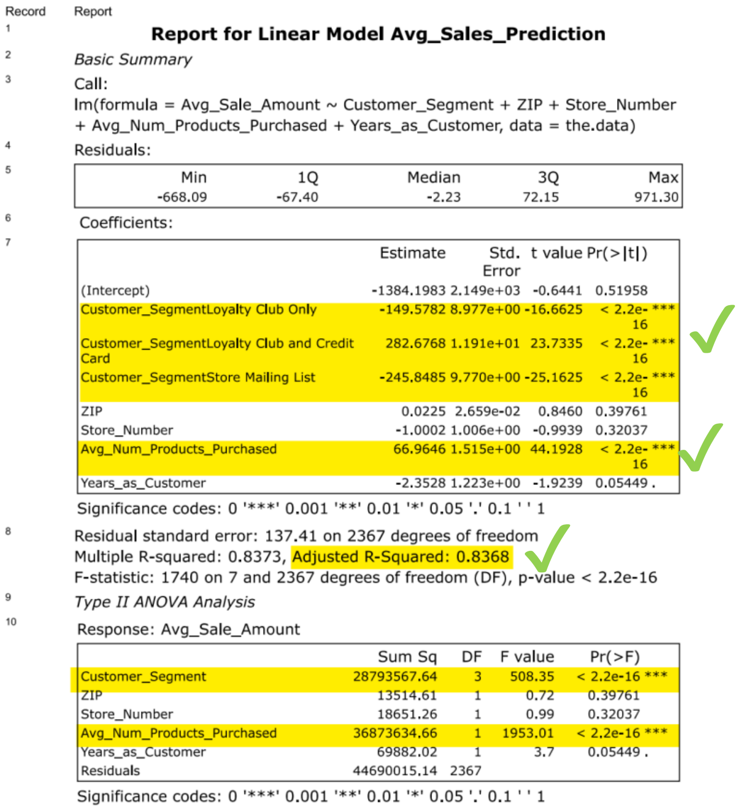
- ☐ Name
- ☑ Customer_Segment
- ☐ Customer_ID
- ☐ Address
- ☐ City
- ☐ State
- ☑ ZIP
- ☐ Avg_Sale_Amount
- ☑ Store_Number
- ☑ Avg_Num_Products_Purchased

Customize

Results - Linear Regression (5) - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages 0 Files All

Last Run Configuration



Record Report

1 **Report for Linear Model Avg_Sales_Prediction**

2 *Basic Summary*

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

4 Residuals:

5

Min	1Q	Median	3Q	Max
-668.09	-67.40	-2.23	72.15	971.30

6 Coefficients:

7

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-1384.1983	2.149e+03	-0.6441	0.51958	
Customer_SegmentLoyalty Club Only	-149.5782	8.977e+00	-16.6625	< 2.2e-16	***
Customer_SegmentLoyalty Club and Credit Card	282.6768	1.191e+01	23.7335	< 2.2e-16	***
Customer_SegmentStore Mailing List	-245.8485	9.770e+00	-25.1625	< 2.2e-16	***
ZIP	0.0225	2.659e-02	0.8460	0.39761	
Store_Number	-1.0002	1.006e+00	-0.9939	0.32037	
Avg_Num_Products_Purchased	66.9646	1.515e+00	44.1928	< 2.2e-16	***
Years_as_Customer	-2.3528	1.223e+00	-1.9239	0.05449	.

✓ ✓

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

8 Residual standard error: 137.41 on 2367 degrees of freedom
Multiple R-squared: 0.8373, Adjusted R-Squared: 0.8368 ✓
F-statistic: 1740 on 7 and 2367 degrees of freedom (DF), p-value < 2.2e-16

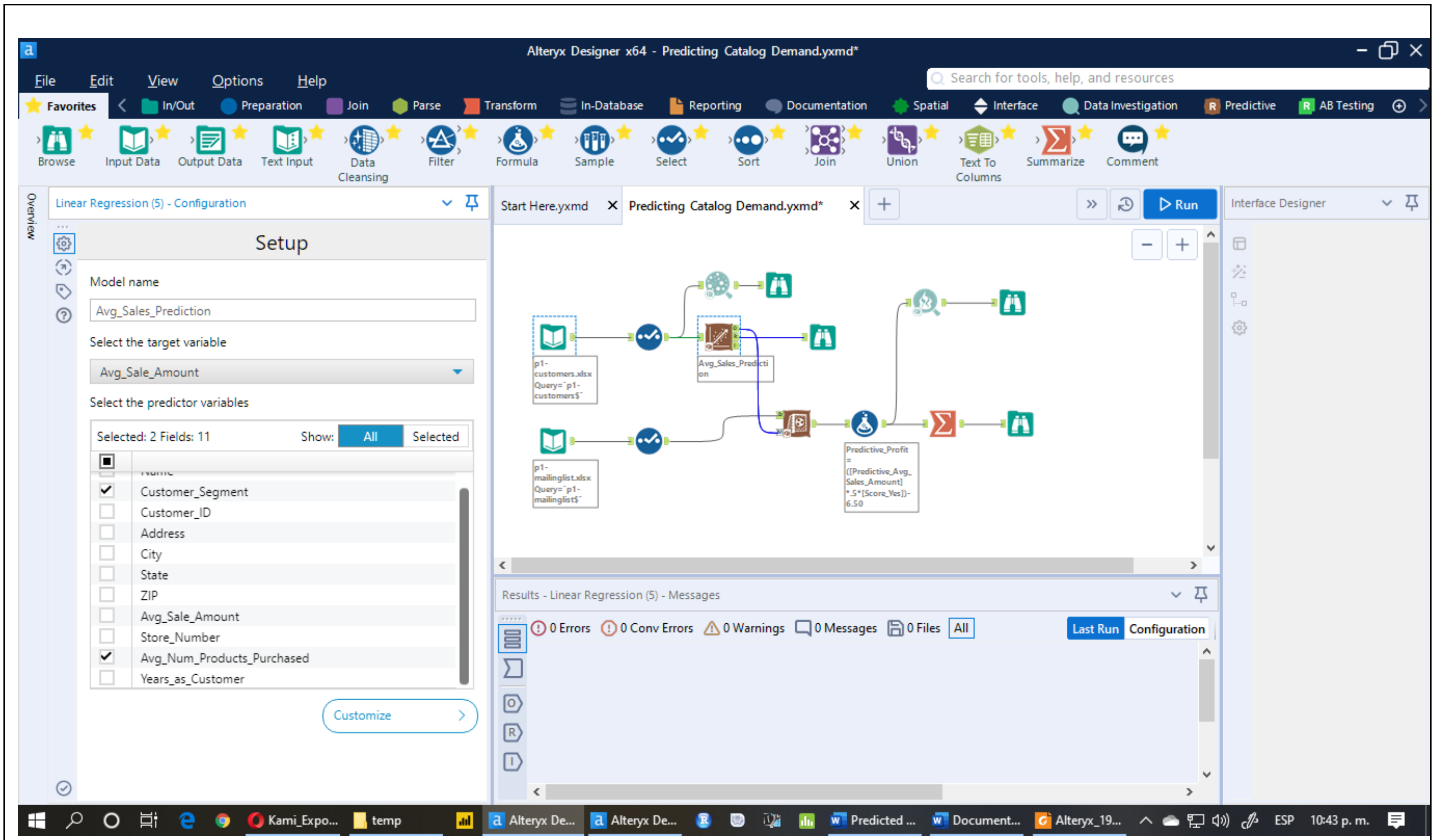
9 *Type II ANOVA Analysis*

10 Response: Avg_Sale_Amount

	Sum Sq	DF	F value	Pr(>F)
Customer_Segment	28793567.64	3	508.35	< 2.2e-16 ***
ZIP	13514.61	1	0.72	0.39761
Store_Number	18651.26	1	0.99	0.32037
Avg_Num_Products_Purchased	36873634.66	1	1953.01	< 2.2e-16 ***
Years_as_Customer	69882.02	1	3.7	0.05449 .
Residuals	44690015.14	2367		

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

2nd. Then I ran the model again, but only with the predictor variables: Customer_Segment and Avg_Num_Products_Purchased, getting:



Alteryx Designer x64 - Predicting Catalog Demand.yxmd*

Linear Regression (5) - Configuration

Setup

Model name

Avg_Sales_Prediction

Select the target variable

Avg_Sale_Amount

Select the predictor variables

Selected: 2 Fields: 11 Show: All Selected

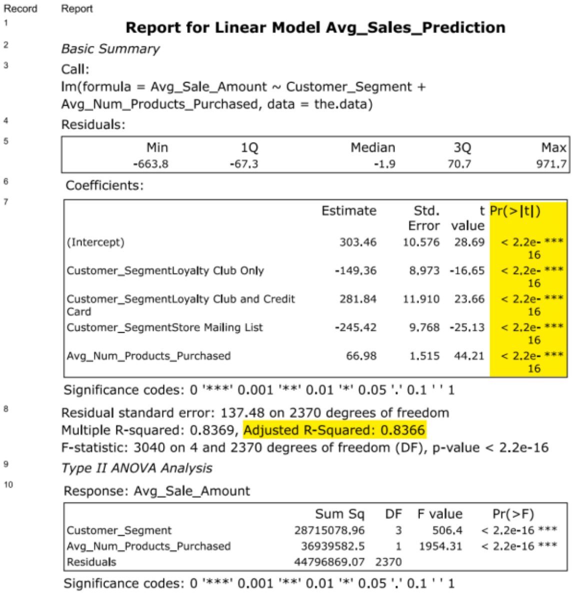
☑ Customer_Segment
□ Customer_ID
□ Address
□ City
□ State
□ ZIP
□ Avg_Sale_Amount
□ Store_Number
☑ Avg_Num_Products_Purchased
□ Years_as_Customer

Customize

Results - Linear Regression (5) - Messages

0 Errors 0 Conv Errors 0 Warnings 0 Messages 0 Files All

File
Edit
View
Options
Help
Search for tools, help, and resources
Favorites
In/Out
Preparation
Join
Parse
Transform
In-Database
Reporting
Documentation
Spatial
Interface
Data Investigation
Predictive
AB Testing
Browse
Input Data
Output Data
Text Input
Data Cleansing
Filter
Formula
Sample
Select
Sort
Join
Union
Text To Columns
Summarize
Comment
Start Here.yxmd
Predicting Catalog Demand.yxmd*
Run
Interface Designer
Overview
p1-customers.xlsx Query=`p1-customers$`
Avg_Sales_Prediction
p1-mailinglist.xlsx Query=`p1-mailinglist$`
Predictive_Profit = ([Predictive_Avg_Sales_Amount] *.5*[Score_Yes])-6.50
Last Run
Configuration
Kami_Expo...
temp
Alteryx De...
Alteryx De...
Predicted ...
Document...
Alteryx_19...
ESP
10:43 p. m.



Record Report

1 **Report for Linear Model Avg_Sales_Prediction**

2 *Basic Summary*

3 Call:
lm(formula = Avg_Sale_Amount ~ Customer_Segment + Avg_Num_Products_Purchased, data = 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 degrees of freedom (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

And with this model build the equation to predict the average sales amount

4. 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)

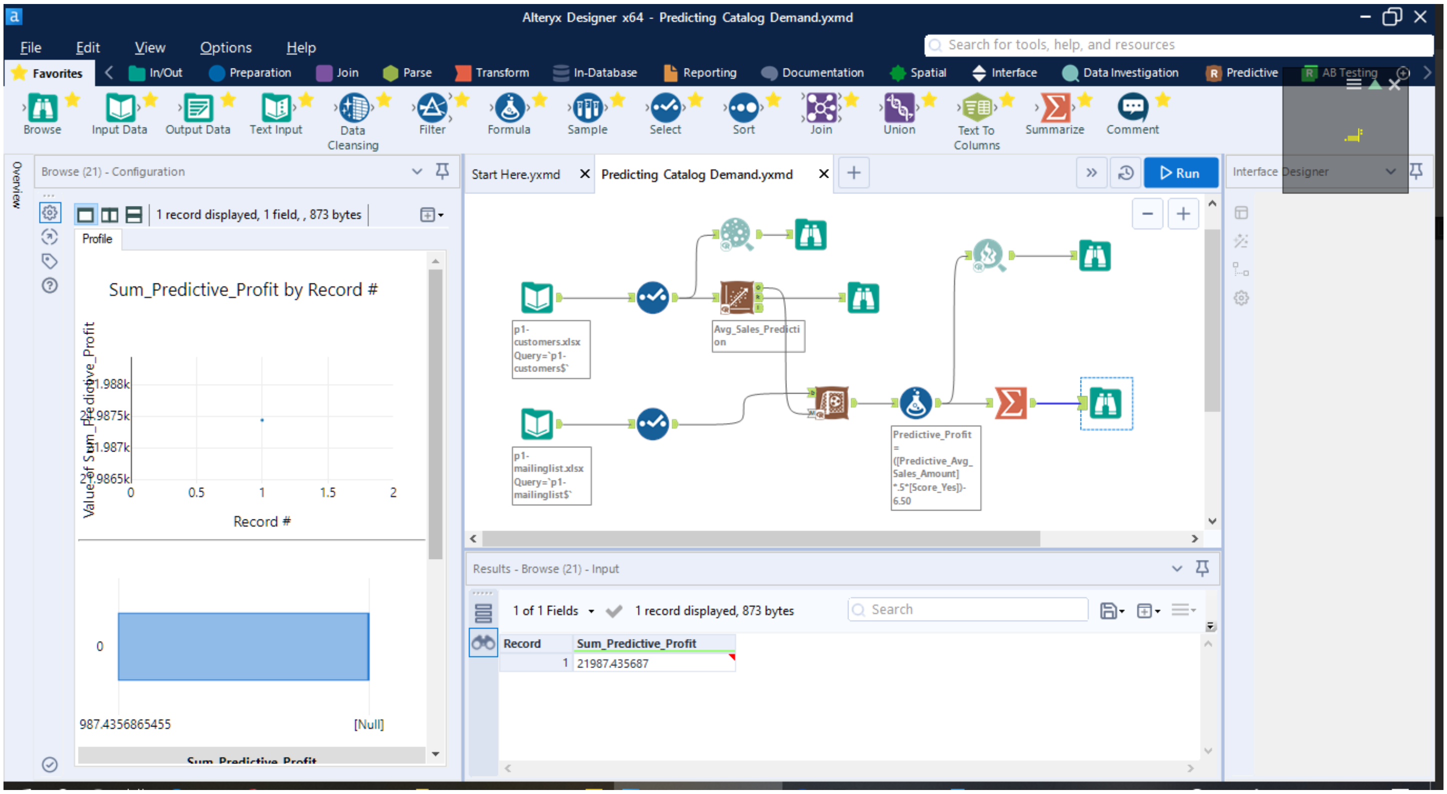
$$Y = 303.46 - 149.36 * Club\ Only + 281.84 * Club\ and\ Credit\ Card - 245.42 * Mailing\ List + 0 * Credit\ Card\ Only + 66.98 * Avg\_Num\_Products\_Purchased$$

Step 3: Presentation/Visualization

1. What is your recommendation? Should the company send the catalog to these 250 customers?
   The company should send the catalog to the 250 new customers, the model shows a good correlation between the predictor variables and the target variable, with a predicted profit over 120% over the profit expected for the manager ($10,000.00)

2. How did you come up with your recommendation?

1st. After looking for the predictor variable that models the target variable according to sadistic analysis, I build the following model in Asterix:



1 record displayed, 1 field, , 873 bytes

p1-customers.xlsx Query=`p1-customers$`

p1-mailinglist.xlsx Query=`p1-mailinglist$`

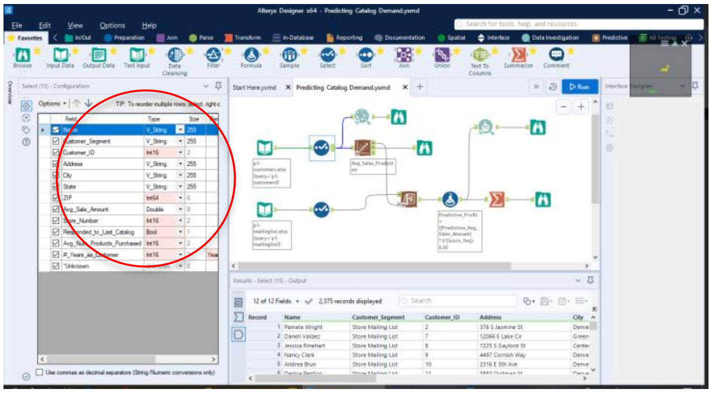
Avg_Sales_Prediction

Predictive_Profit = ([Predictive_Avg_Sales_Amount] *.5*[Score_Yes])-6.50

1 record displayed, 873 bytes

Record	Sum_Predictive_Profit
1	21987.435687

I made some consideration with the raw data as is shown:



Alteryx Designer x64 - Predicting Catalog Demand.yxmd

Select (15) - Configuration

TIP: To reorder multiple rows: select, right-c

	Field	Type	Size
☑	Name	V_String	255
☑	Customer_Segment	V_String	255
☑	Customer_ID	Int16	2
☑	Address	V_String	255
☑	City	V_String	255
☑	State	V_String	255
☑	ZIP	Int64	8
☑	Avg_Sale_Amount	Double	8
☑	Store_Number	Int16	2
☑	Responded_to_Last_Catalog	Bool	1
☑	Avg_Num_Products_Purchased	Int16	2
☑	#_Years_as_Customer	Int16	2
☑	*Unknown	Unknown	0

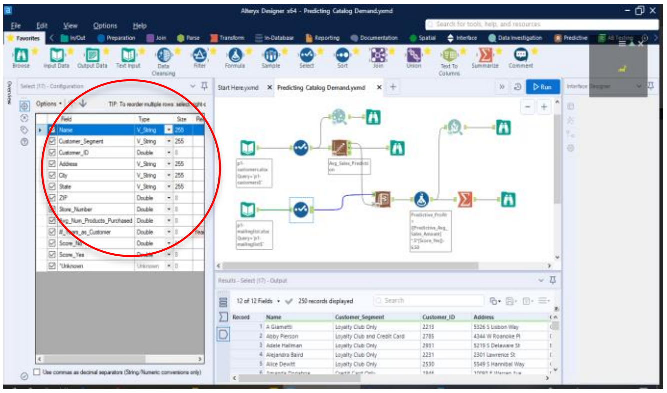
□ Use commas as decimal separators (String/Numeric conversions only)

Predictive_Profit = ([Predictive_Avg_Sales_Amount]*.5*[Score_Yes])-6.50

Results - Select (15) - Output

12 of 12 Fields ✓ 2,375 records displayed

Record	Name	Customer_Segment	Customer_ID	Address	City
1	Pamela Wright	Store Mailing List	2	376 S Jasmine St	Denve
2	Danell Valdez	Store Mailing List	7	12066 E Lake Cir	Green
3	Jessica Rinehart	Store Mailing List	8	7225 S Gaylord St	Center
4	Nancy Clark	Store Mailing List	9	4497 Cornish Way	Denve
5	Andrea Brun	Store Mailing List	10	2316 E 5th Ave	Denve



Alteryx Designer x64 - Predicting Catalog Demand.yxmd

Select (17) - Configuration

	Field	Type	Size
☑	Name	V_String	255
☑	Customer_Segment	V_String	255
☑	Customer_ID	Double	8
☑	Address	V_String	255
☑	City	V_String	255
☑	State	V_String	255
☑	ZIP	Double	8
☑	Store_Number	Double	8
☑	Avg_Num_Products_Purchased	Double	8
☑	#_Years_as_Customer	Double	8
☑	Score_No	Double	8
☑	Score_Yes	Double	8
☑	*Unknown	Unknown	0

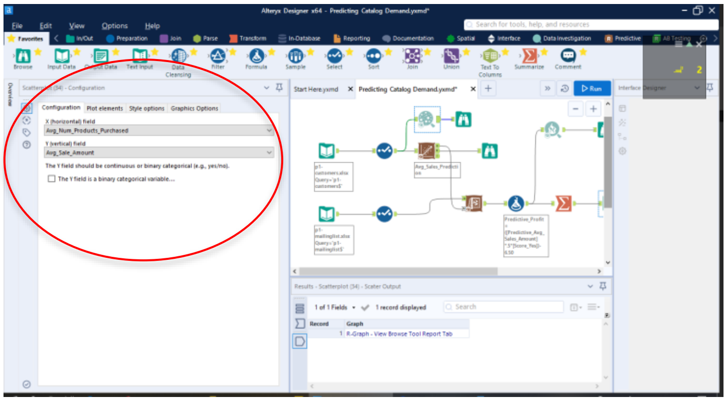
□ Use commas as decimal separators (String/Numeric conversions only)

Results - Select (17) - Output

12 of 12 Fields ✓ 250 records displayed

Record	Name	Customer_Segment	Customer_ID	Address
1	A Giametti	Loyalty Club Only	2213	5326 S Lisbon Way
2	Abby Pierson	Loyalty Club and Credit Card	2785	4344 W Roanoke Pl
3	Adele Hallman	Loyalty Club Only	2931	5219 S Delaware St
4	Alejandra Baird	Loyalty Club Only	2231	2301 Lawrence St
5	Alice Dewitt	Loyalty Club Only	2530	5549 S Hannibal Way

2nd. I generated a scatter plot with the following configuration (it helps me to see the correlation between predictor and target variable):



Scatterplot (34) - Configuration

Configuration | Plot elements | Style options | Graphics Options

X (horizontal) field: Avg_Num_Products_Purchased

Y (vertical) field: Avg_Sale_Amount

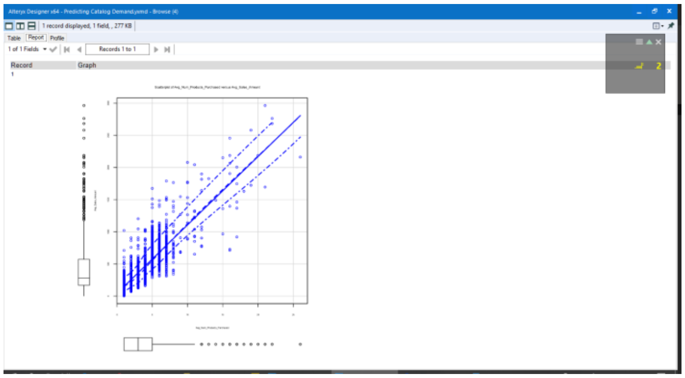
The Y field should be continuous or binary categorical (e.g., yes/no).

□ The Y field is a binary categorical variable...

Results - Scatterplot (34) - Scater Output

1 of 1 Fields ✓ 1 record displayed

Record	Graph
1	R-Graph - View Browse Tool Report Tab



Alteryx Designer x64 - Predicting Catalog Demand.yxmd - Browse (4)

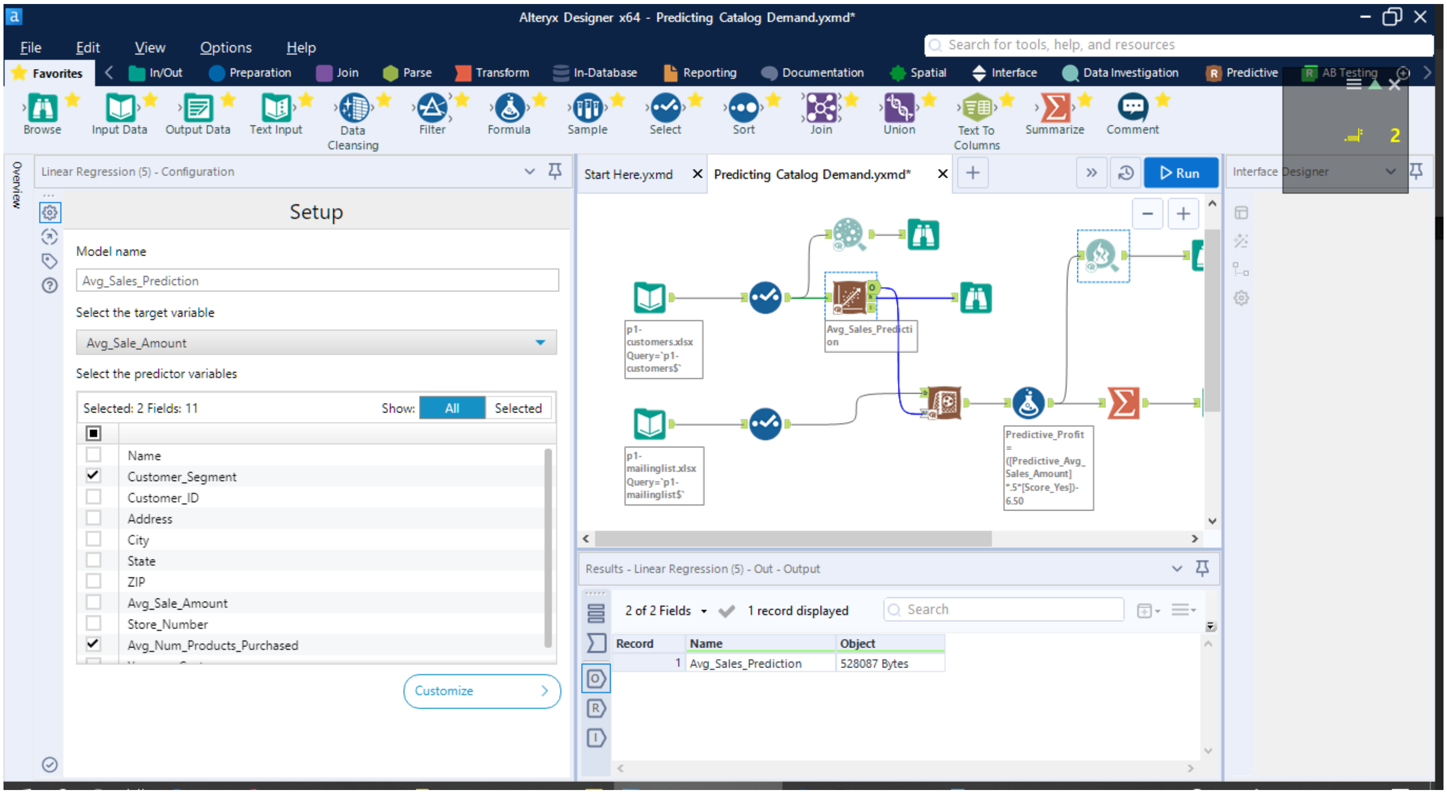
1 record displayed, 1 field, , 277 KB

Table | Report | Profile

1 of 1 Fields ✓ Records 1 to 1

Record	Graph
1	

3rd. I build the linear regression as is shown:



Alteryx Designer x64 - Predicting Catalog Demand.yxmd*

Linear Regression (5) - Configuration

Setup

Model name: Avg_Sales_Prediction

Select the target variable: Avg_Sale_Amount

Select the predictor variables

Selected: 2 Fields: 11 Show: All Selected

□ Name
☑ Customer_Segment
□ Customer_ID
□ Address
□ City
□ State
□ ZIP
□ Avg_Sale_Amount
□ Store_Number
☑ Avg_Num_Products_Purchased

Customize

p1-customers.xlsx Query=`p1-customers$`

p1-mailinglist.xlsx Query=`p1-mailinglist$`

Avg_Sales_Prediction

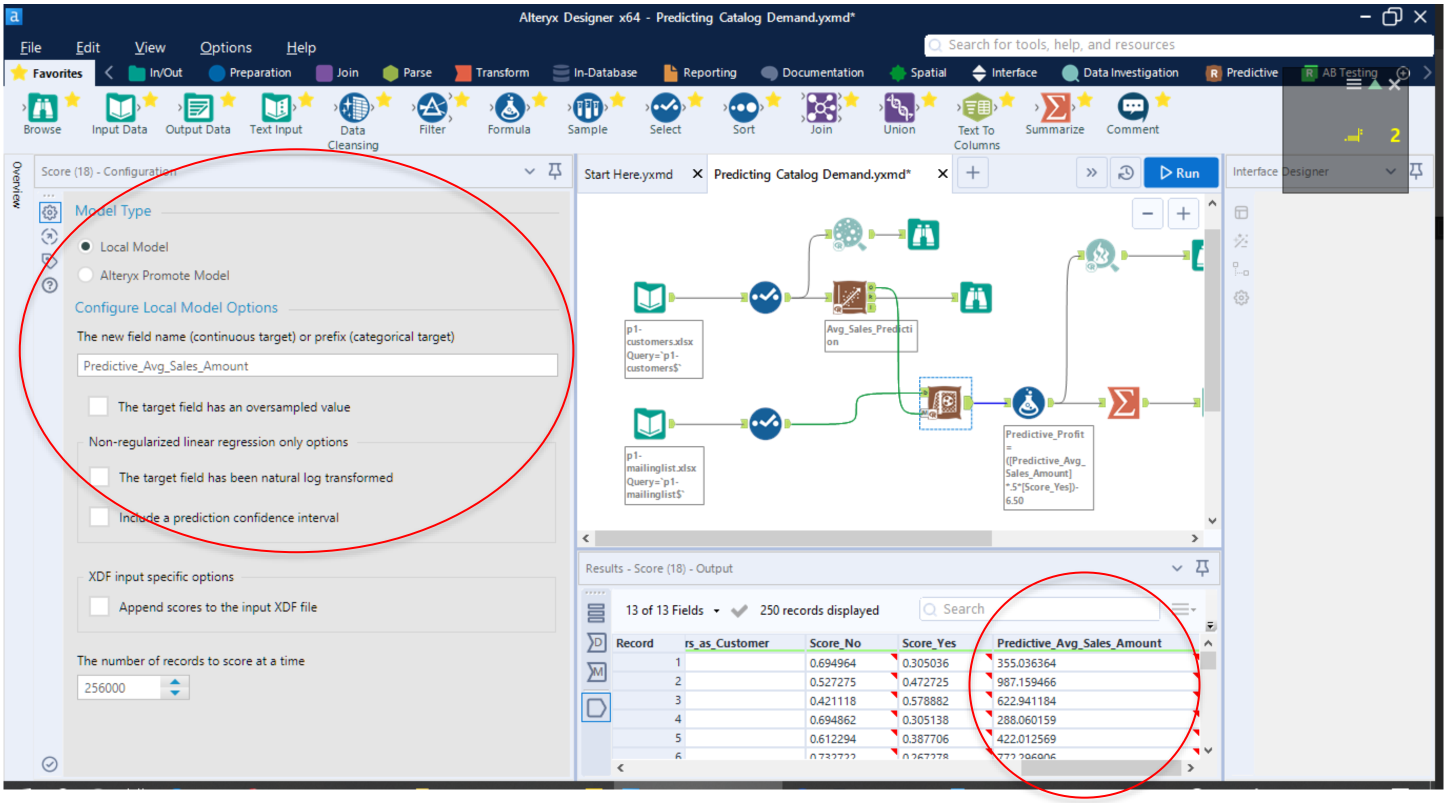
Predictive_Profit = ([Predictive_Avg_Sales_Amount]*.5*[Score_Yes])-6.50

Results - Linear Regression (5) - Out - Output

2 of 2 Fields ✓ 1 record displayed

Record	Name	Object
1	Avg_Sales_Prediction	528087 Bytes

4th. Using the Avg-Sales Prediction from the linear regression tool and the scoring tool I aggregate the column Predicted_Avg_Sales_amount for the 250 new customers as is it shows:



Score (18) - Configuration

Model Type

● Local Model

○ Alteryx Promote Model

Configure Local Model Options

The new field name (continuous target) or prefix (categorical target)

Predictive_Avg_Sales_Amount

□ The target field has an oversampled value

Non-regularized linear regression only options

□ The target field has been natural log transformed

□ Include a prediction confidence interval

XDF input specific options

□ Append scores to the input XDF file

The number of records to score at a time

256000

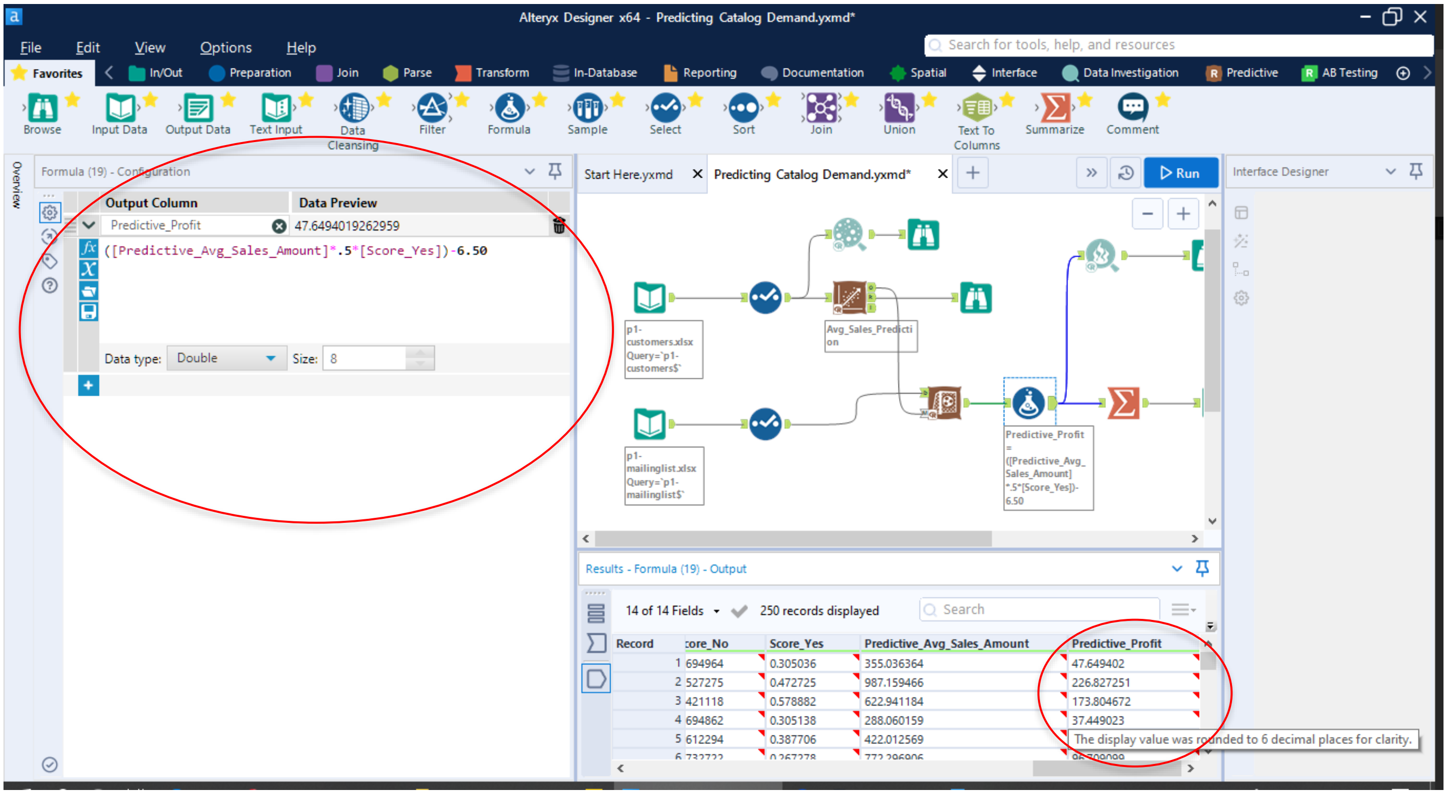
Results - Score (18) - Output

13 of 13 Fields ✓ 250 records displayed

Record	rs_as_Customer	Score_No	Score_Yes	Predictive_Avg_Sales_Amount
1		0.694964	0.305036	355.036364
2		0.527275	0.472725	987.159466
3		0.421118	0.578882	622.941184
4		0.694862	0.305138	288.060159
5		0.612294	0.387706	422.012569
6		0.732722	0.267278	772.296906

5th. Then I calculated the predicted profit by ever customer using the formula:

$$Predicted\_Proficit = (Predictive\_Avg\_Sales\_Amount * 0.5 * Score\_Yes) - 6.50$$



Formula (19) - Configuration

Output Column	Data Preview
Predictive_Profit	47.6494019262959

```
([Predictive_Avg_Sales_Amount]*.5*[Score_Yes])-6.50
```

Data type: Double Size: 8

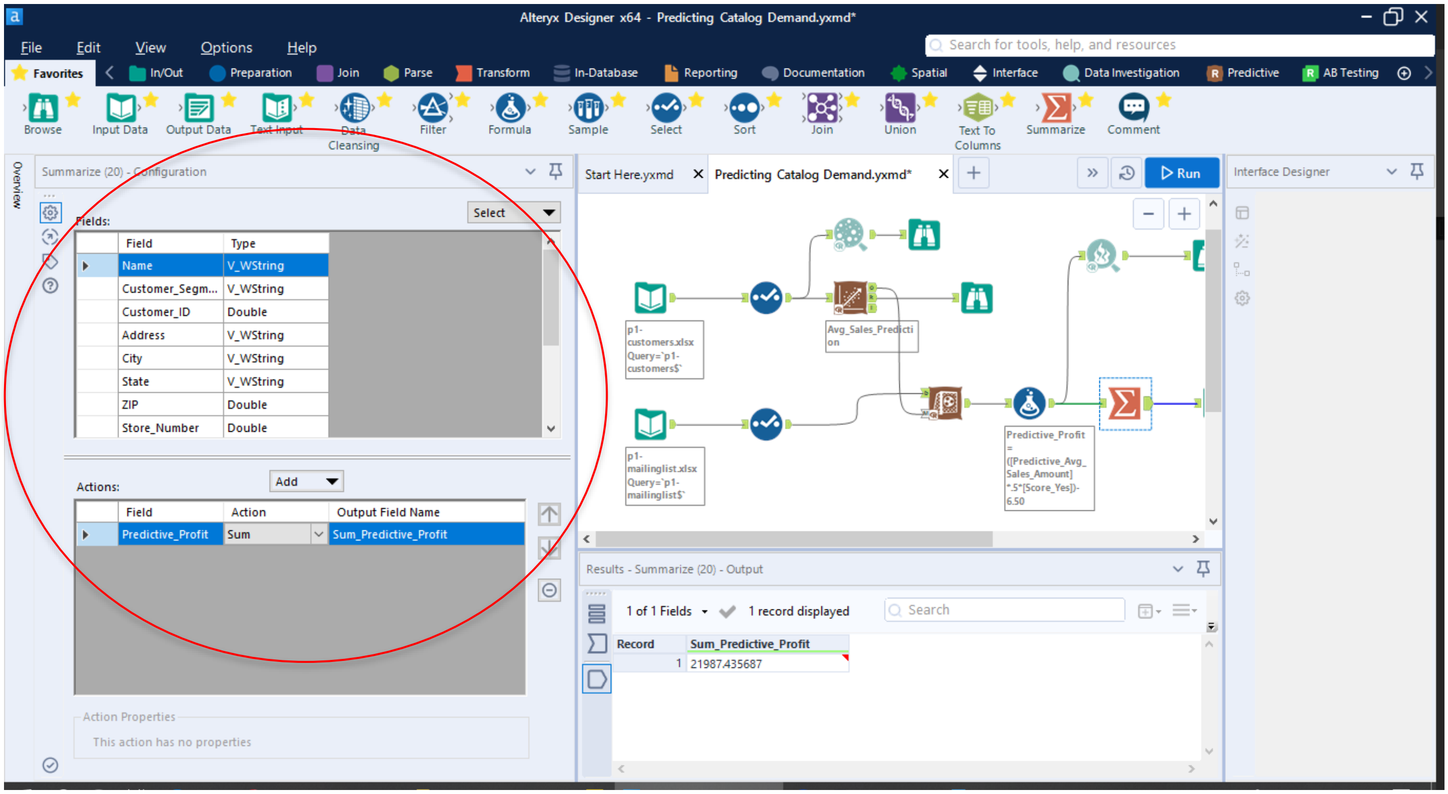
Results - Formula (19) - Output

14 of 14 Fields ✓ 250 records displayed

Record	core_No	Score_Yes	Predictive_Avg_Sales_Amount	Predictive_Profit
1	694964	0.305036	355.036364	47.649402
2	527275	0.472725	987.159466	226.827251
3	421118	0.578882	622.941184	173.804672
4	694862	0.305138	288.060159	37.449023
5	612294	0.387706	422.012569	
6	732722	0.267278	772.296906	96.709099

The display value was rounded to 6 decimal places for clarity.

6th. Finally, I summary the predicted profit by the 250 new customers



Summarize (20) - Configuration

Fields: Select

Field	Type
Name	V_WString
Customer_Segm...	V_WString
Customer_ID	Double
Address	V_WString
City	V_WString
State	V_WString
ZIP	Double
Store_Number	Double

Actions: Add

Field	Action	Output Field Name
Predictive_Profit	Sum	Sum_Predictive_Profit

Action Properties

This action has no properties

p1-customers.xlsx Query='p1-customers$'

Avg_Sales_Prediction

p1-mailinglist.xlsx Query='p1-mailinglist$'

Predictive_Profit = ([Predictive_Avg_Sales_Amount]*.5*[Score_Yes])-6.50

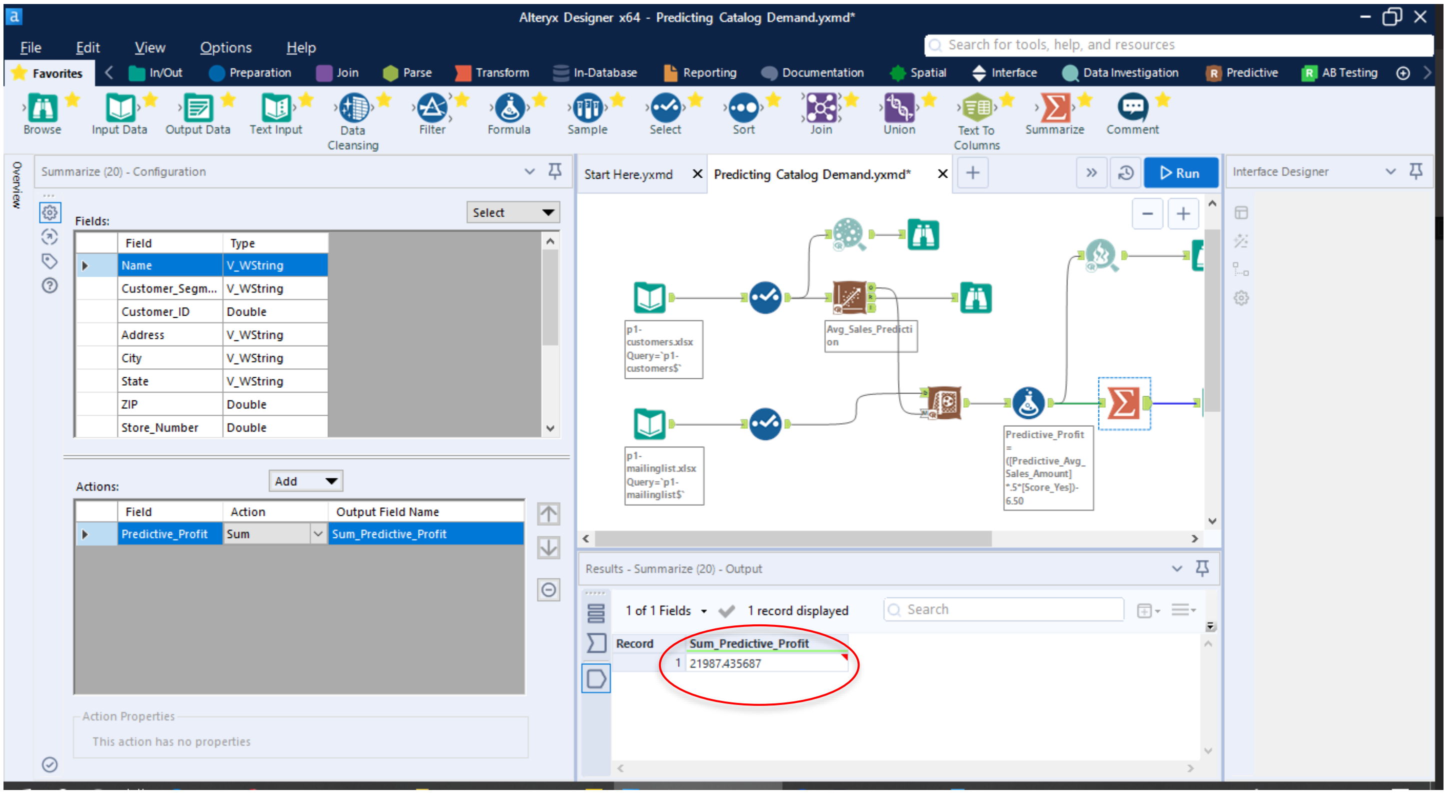
Results - Summarize (20) - Output

1 of 1 Fields ✓ 1 record displayed Search

Record	Sum_Predictive_Profit
1	21987.435687

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

As is shown in the calculus the expected profit will be $21,379.1 114$ over the manager profit expected $10,000.00



Summarize (20) - Configuration

Fields: Select

Field	Type
Name	V_WString
Customer_Segm...	V_WString
Customer_ID	Double
Address	V_WString
City	V_WString
State	V_WString
ZIP	Double
Store_Number	Double

Actions: Add

Field	Action	Output Field Name
Predictive_Profit	Sum	Sum_Predictive_Profit

Action Properties

This action has no properties

p1-customers.xlsx Query='p1-customers$'

Avg_Sales_Prediction

p1-mailinglist.xlsx Query='p1-mailinglist$'

Predictive_Profit = ([Predictive_Avg_Sales_Amount]*.5*[Score_Yes])-6.50

Results - Summarize (20) - Output

1 of 1 Fields ✓ 1 record displayed Search

Record	Sum_Predictive_Profit
1	21987.435687