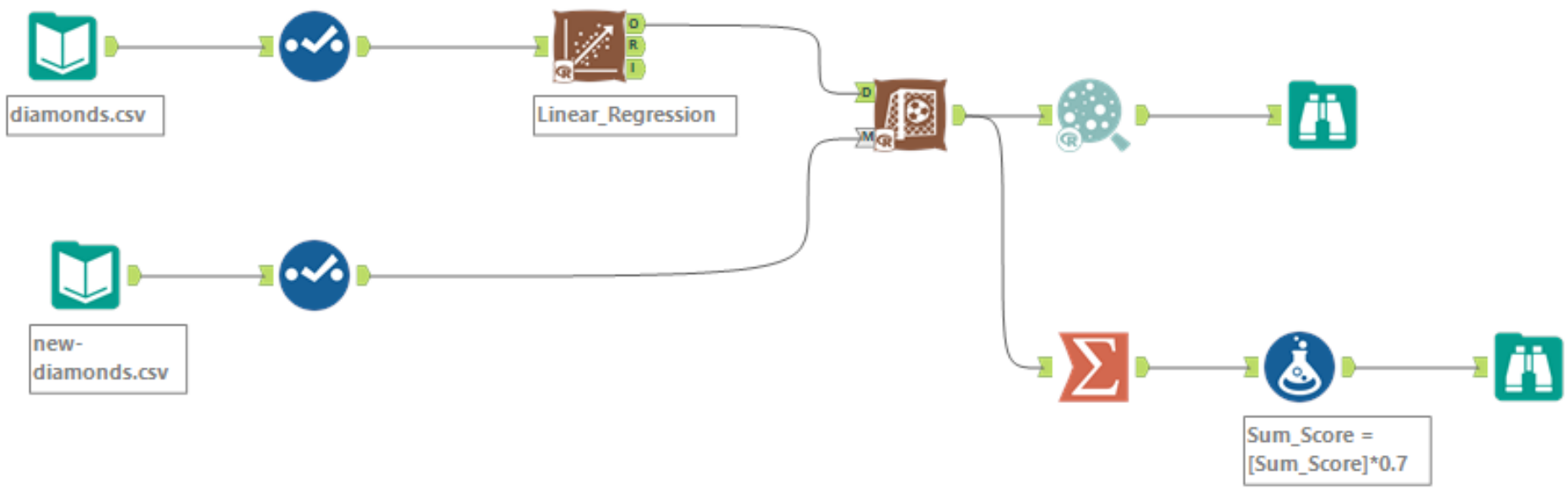
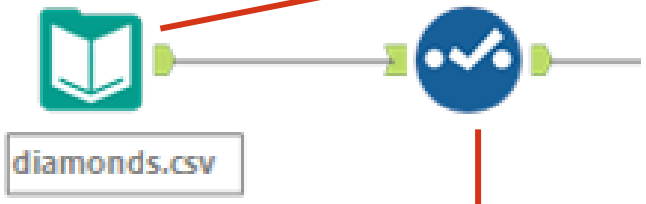
DIAMOND PRICES ANALYSIS

Using Alteryx

WORKFLOW


diamonds.csv
Linear_Regression
new-
diamonds.csv
Sum_Score =
[Sum_Score]*0.7

STEP 1: TRAIN DATA



	Field	Type	Size	Rename	Description
▶	Field_1	V_String	254		
	carat	Double	8		
	cut	V_String	254		
	cut_ord	Double	8		
	color	V_String	254		
	clarity	V_String	254		
	clarity_ord	Double	8		
	price	Double	8		
	*Unknown	Unknown	0		Dynamic or Unknown Fields

Connect a File or Database

C:\Users\Hafizah\Desktop\Alteryx\Lesson 4\diamonds.csv

Options

	Name	Value
1	Record Limit	
2	File Format	Comma Separated Value (*.csv)
3	Search SubDirs	☐
4	Output File Name as Field	No
5	Delimiters	,
6	First Row Contains Field Names	☑

Preview (first 100 records) Refresh

	Field_1	carat	cut	cut_ord	color	clarity	clarity_ord	price
1	1	0.51	Premium	4	F	VS1	4	1749
2	2	2.25	Fair	1	G	I1	1	7069
3	3	0.7	Very Good	3	E	VS2	5	2757
4	4	0.47	Good	2	F	VS1	4	1243
5	5	0.3	Ideal	5	G	VVS1	7	789
6	6	0.33	Ideal	5	D	SI1	3	728
7	7	2.01	Very Good	3	G	SI1	3	18398
8	8	0.51	Ideal	5	F	VVS2	6	2203
9	9	1.7	Premium	4	D	SI1	3	15100
10	10	0.53	Premium	4	D	VS2	5	1857
11	11	0.39	Premium	4	H	SI1	3	834
12	12	1.5	Very Good	3	H	SI1	3	7708
13	13	1	Premium	4	E	VS2	5	6272
14	14	1.29	Ideal	5	J	VS1	4	5676
15	15	2.01	Good	2	D	SI2	2	16776
16	16	1.13	Ideal	5	G	VS1	4	7404
17	17	0.7	Ideal	5	I	SI2	2	1702
18	18	0.38	Very Good	3	I	VS1	4	606
19	19	1.17	Ideal	5	H	SI2	2	5423
20	20	1.51	Premium	4	F	SI1	3	8033
21	21	0.4	Ideal	5	D	VVS1	7	1279
22	22	0.41	Very Good	3	F	VS2	5	863
23	23	0.51	Ideal	5	G	VVS1	7	1893
24	24	1	Premium	4	H	SI2	2	3584
25	25	1.09	Ideal	5	F	VVS2	6	10196
26	26	0.39	Good	2	E	VS1	4	1082
27	27	0.72	Premium	4	E	VS2	5	3024
28	28	1.14	Very Good	3	E	SI2	2	5593
29	29	0.3	Ideal	5	D	VS2	5	710

STEP 2: MODEL

Model name

Linear_Regression

Select the target variable

price

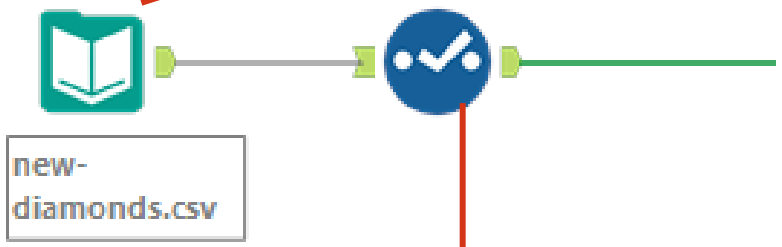
Select the predictor variables

Selected: 4 Fields: 7 Show: All Selected

☐	Field_1
☑	carat
☐	cut
☑	cut_ord
☑	color
☐	clarity
☑	clarity_ord

Customize

STEP 3 : TEST DATA



	Field	Type	Size	Rename	Description
▶	Field_1	V_String	254		
	carat	Double	8		
	cut	V_String	254		
	cut_ord	Double	8		
	color	V_String	254		
	clarity	V_String	254		
	clarity_ord	Double	8		
	*Unknown	Unknown	0		Dynamic or Unknown Fields

Input Data (6) - Configuration

Connect a File or Database

C:\Users\Hafizah\Desktop\Alteryx\Lesson 4\new-diamonds.csv

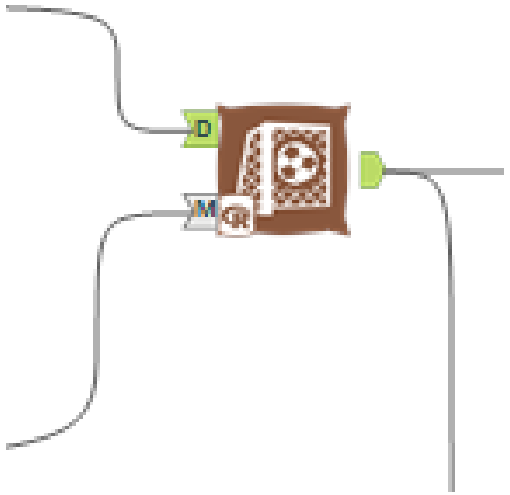
Options

	Name	Value
1	Record Limit	
2	File Format	Comma Separated Value (*.csv)
3	Search SubDirs	☐
4	Output File Name as Field	No
5	Delimiters	,
6	First Row Contains Field Names	☑

Preview (first 100 records)

	Field_1	carat	cut	cut_ord	color	clarity	clarity_ord
1	1	1.22	Premium	4	G	SI1	3
2	2	1.01	Good	2	G	VS2	5
3	3	0.71	Very Good	3	I	VS2	5
4	4	1.01	Ideal	5	D	SI2	2
5	5	0.27	Ideal	5	H	VVS2	6
6	6	0.52	Premium	4	G	VS1	4
7	7	1.01	Premium	4	F	SI1	3
8	8	0.59	Ideal	5	D	SI1	3
9	9	1.01	Good	2	E	SI1	3
10	10	2.03	Ideal	5	F	SI2	2
11	11	1.35	Premium	4	H	VS2	5
12	12	0.74	Ideal	5	G	SI1	3
13	13	0.9	Premium	4	D	SI1	3
14	14	0.3	Good	2	G	VS2	5
15	15	1.01	Good	2	F	VS2	5
16	16	1.02	Good	2	H	SI2	2
17	17	2.05	Premium	4	G	SI1	3
18	18	0.54	Ideal	5	I	SI1	3
19	19	0.72	Ideal	5	G	VS2	5
20	20	2	Premium	4	J	SI2	2
21	21	1.57	Premium	4	G	SI2	2
22	22	0.89	Premium	4	G	SI1	3
23	23	0.33	Premium	4	I	VVS2	6
24	24	0.3	Very Good	3	G	IF	8
25	25	1.79	Very Good	3	I	VS2	5
26	26	1.11	Ideal	5	E	SI2	2
27	27	0.79	Premium	4	F	SI1	3
28	28	0.71	Very Good	3	F	SI1	3
29	29	0.73	Ideal	5	G	SI1	3

STEP 4: SCORE TOOL



Model Type

- (•) Local Model
- ( ) Alteryx Promote Model

Configure Local Model Options

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

[ ] 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

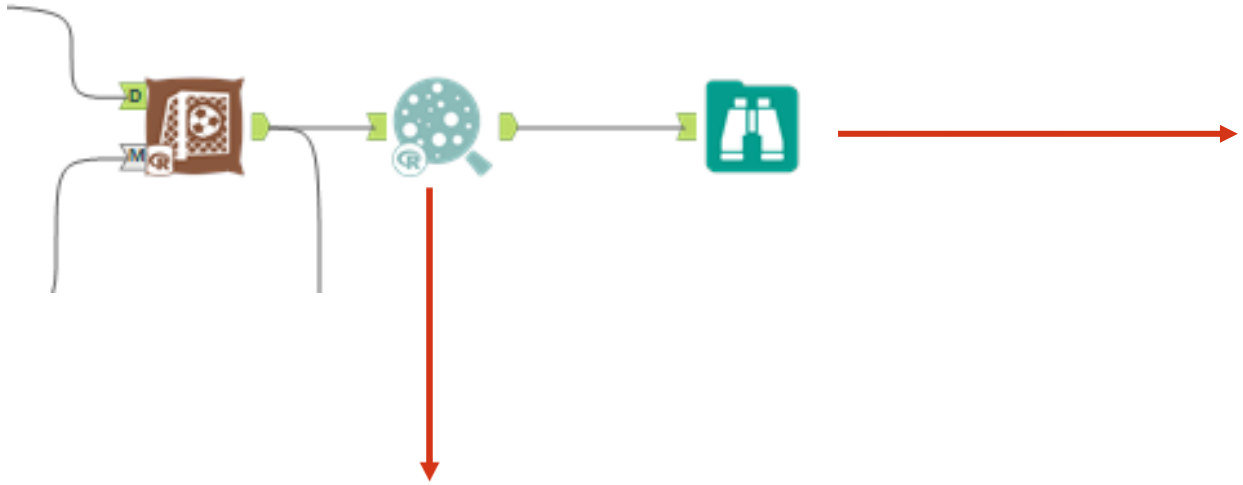
The number of records to score at a time

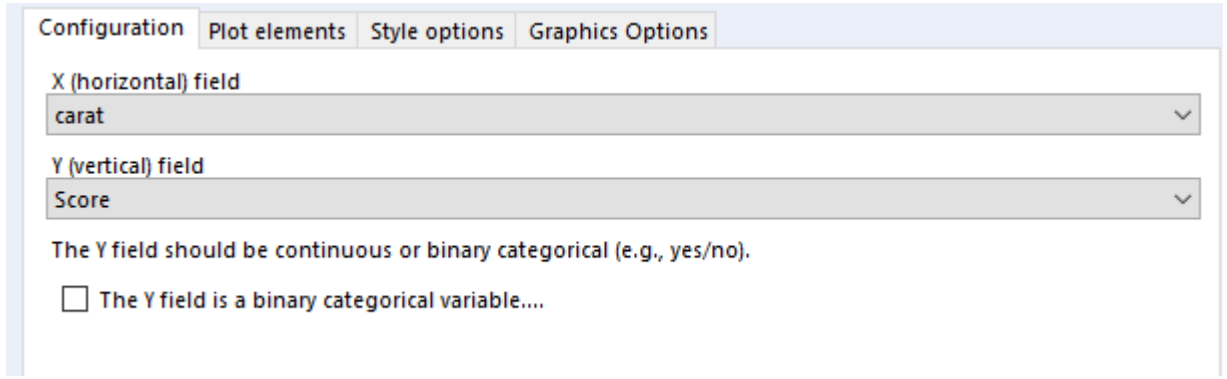
256000

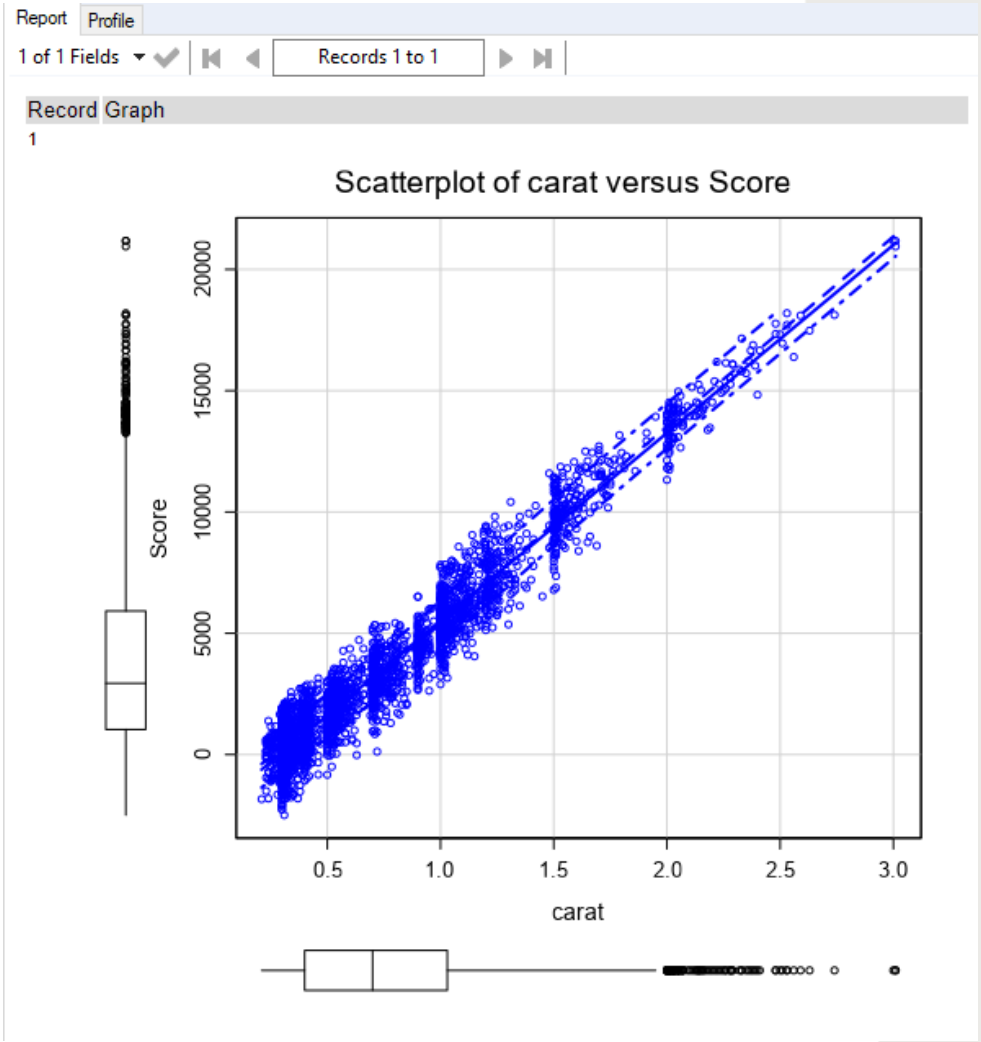
8 of 8 Fields | Cell Viewer | 3,000 records displayed | Search

Record	Field_1	carat	cut	cut_ord	color	clarity	clarity_ord	Score
1	1	1.22	Premium	4	G	SI1	3	7,213.018564
2	2	1.01	Good	2	G	VS2	5	6,027.359861
3	3	0.71	Very Good	3	I	VS2	5	2,659.717162
4	4	1.01	Ideal	5	D	SI2	2	5,526.012885
5	5	0.27	Ideal	5	H	VVS2	6	32.230452
6	6	0.52	Premium	4	G	VS1	4	1,572.784831
7	7	1.01	Premium	4	F	SI1	3	5,544.914185
8	8	0.59	Ideal	5	D	SI1	3	2,335.690656

STEP 5: SCATTER PLOT







STEP 6: SUM & MARGIN