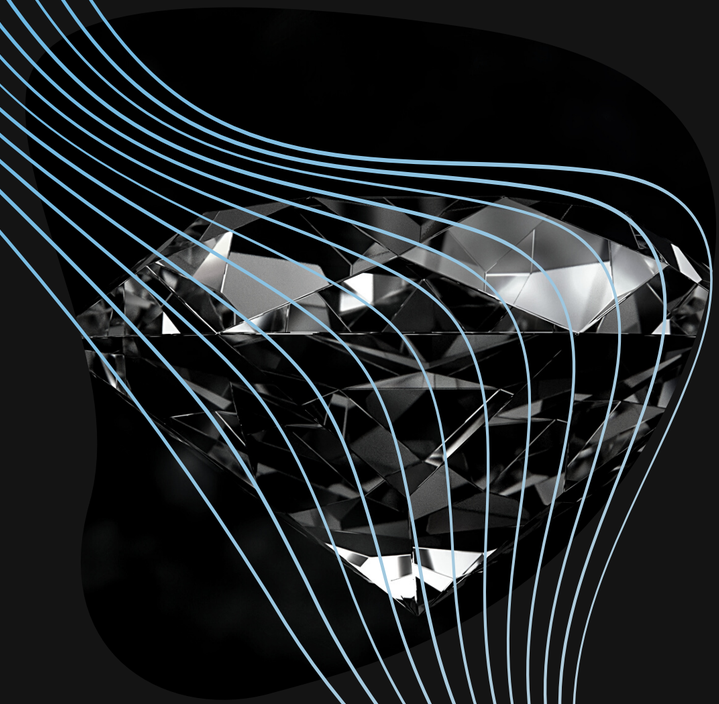
Predicting Diamond Prices

PREDICTIVE ANALYTICS FOR BUSINESS

Table of Contents



Business Problem

"A diamond distributor has recently decided to exit the market and has put up a set of 3,000 diamonds up for auction. Seeing this as a great opportunity to expand its inventory, a jewelry company has shown interest in making a bid. You, as the business analysts, are tasked to apply that model to make a recommendation for how much the company should bid for the entire set of 3,000 diamonds.

Note: The diamond price that the model predicts represents the final retail price the consumer will pay. The company generally purchases diamonds from distributors at 70% of that price, so your recommended bid price should represent that."

Data Understanding

2 data sources:

- diamonds.cvs
- newdiamonds.cvs

Data Characteristics:

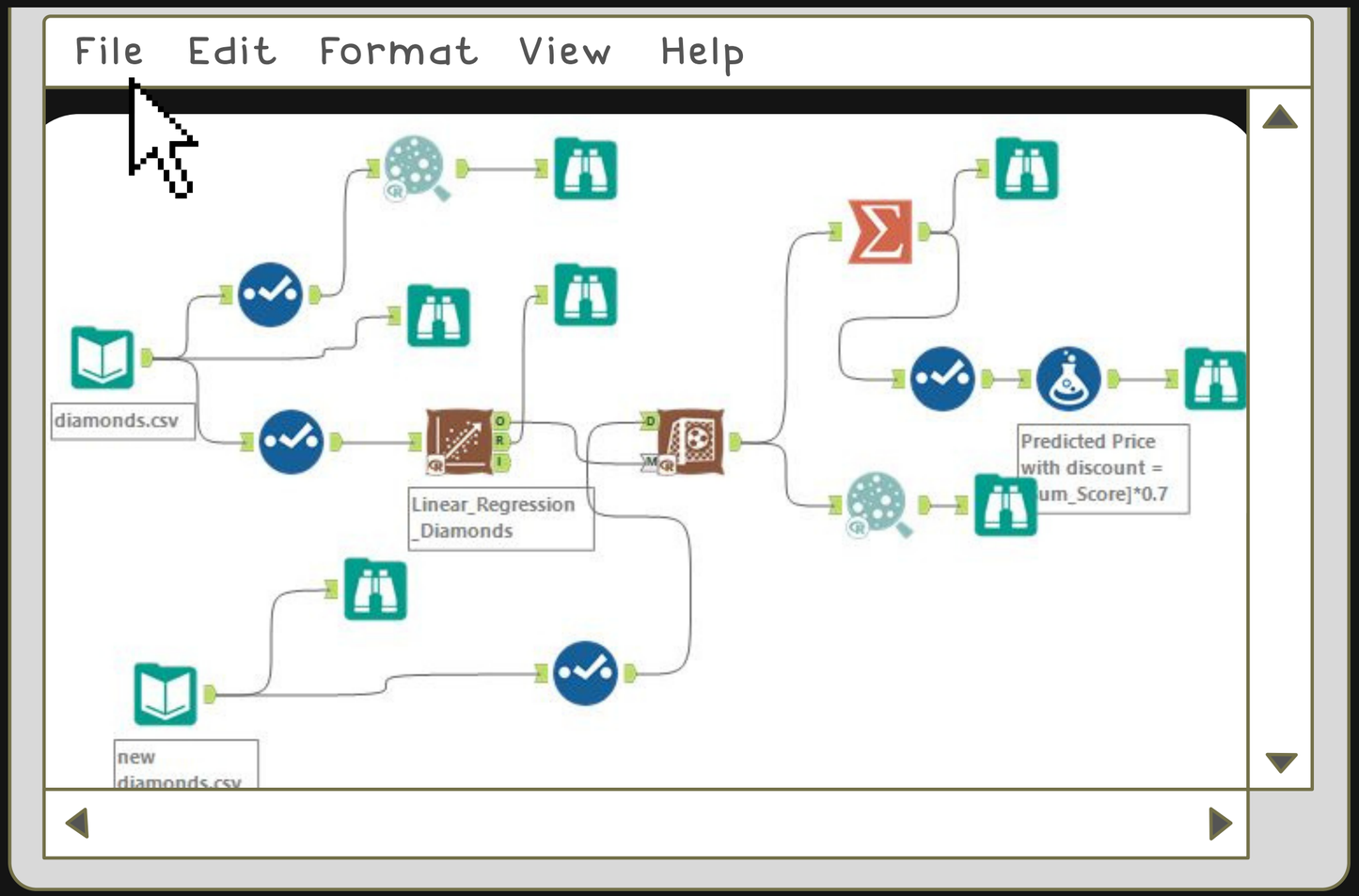
- First dataset is a collection of 50,000 diamonds with data on cut, clarity, color, carat weight, and retail price.
- Second dataset contains the data (cut, clarity, color, carat weight) for the batch (3,000) the company would like to purchase.

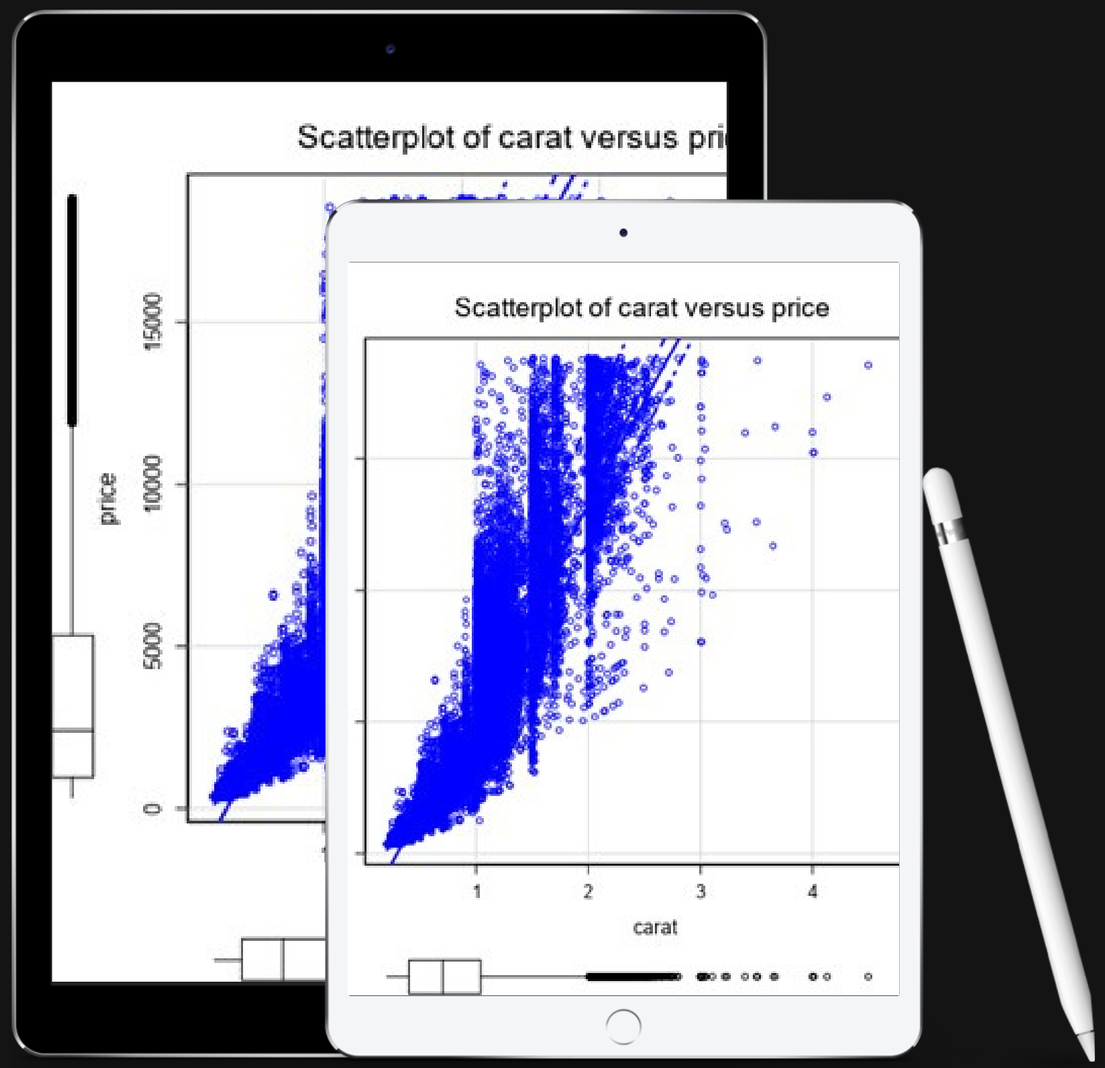
Data-rich scenario with both categorical and numeric variables.

Applied Method:
Linear Regression with Alteryx.

Analysis and Modeling

WITH ALTERYX





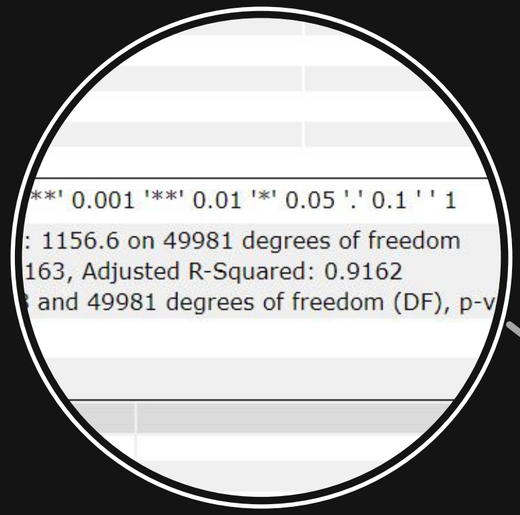
Analysis and Modeling

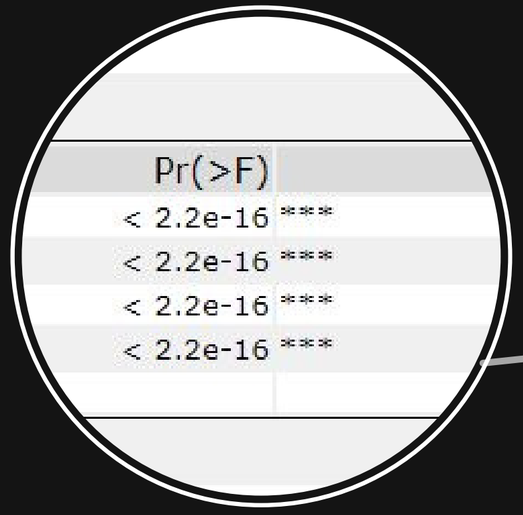
When browsing our diamonds.cvs database, the scatterplot tool indicates a linear relationship between carat weight and diamond price.

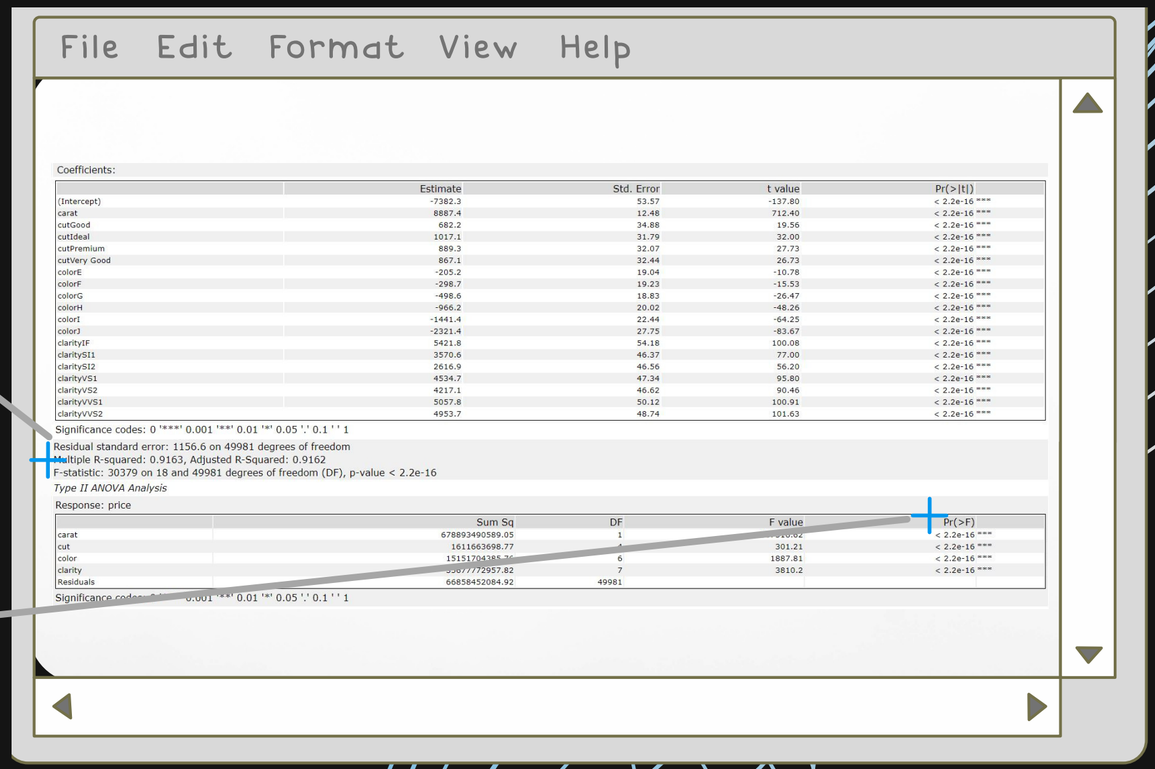
Let's now plug in the available predictor variables to predict diamond prices in a more accurate way.

Analysis and Modeling

The predictors are significant variables (at the significance level of .001.) for predicting the diamond prices, as shown by the significance code '***', next to the p-value of the variable.
Moreover, the adjusted R-squared Value (0.9162) suggests that nearly all variance in the target variable can be explained by our model.

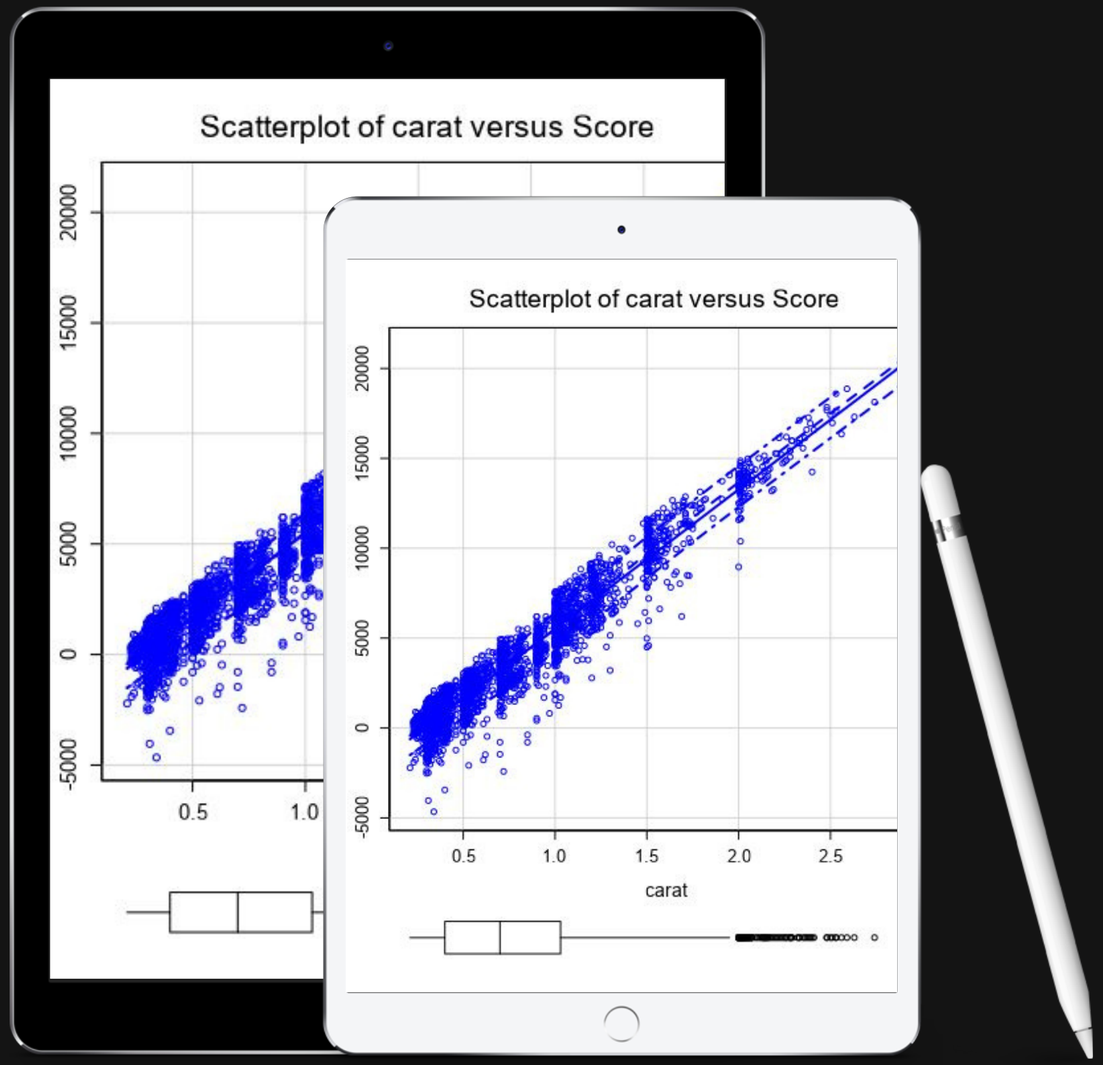






	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-7382.3	53.57	-137.80	< 2.2e-16 ***
carat	8887.4	12.48	712.40	< 2.2e-16 ***
cutGood	682.2	34.88	19.56	< 2.2e-16 ***
cutIdeal	1017.1	31.79	32.00	< 2.2e-16 ***
cutPremium	889.3	32.07	27.73	< 2.2e-16 ***
cutVery Good	867.1	32.44	26.73	< 2.2e-16 ***
colorE	-205.2	19.04	-10.78	< 2.2e-16 ***
colorF	-298.7	19.23	-15.53	< 2.2e-16 ***
colorG	-498.6	18.83	-26.47	< 2.2e-16 ***
colorH	-966.2	20.02	-48.26	< 2.2e-16 ***
colorI	-1441.4	22.44	-64.25	< 2.2e-16 ***
colorJ	-2321.4	27.75	-83.67	< 2.2e-16 ***
clarityIF	5421.8	54.18	100.08	< 2.2e-16 ***
claritySI1	3570.6	46.37	77.00	< 2.2e-16 ***
claritySI2	2616.9	46.56	56.20	< 2.2e-16 ***
clarityVS1	4534.7	47.34	95.80	< 2.2e-16 ***
clarityVS2	4217.1	46.62	90.46	< 2.2e-16 ***
clarityVVS1	5057.8	50.12	100.91	< 2.2e-16 ***
clarityVVS2	4953.7	48.74	101.63	< 2.2e-16 ***

	Sum Sq	DF	F value	Pr(>F)
carat	678893490589.05	1	[illegible]	< 2.2e-16 ***
cut	1611663698.77	4	301.21	< 2.2e-16 ***
color	15151704385.76	6	1887.81	< 2.2e-16 ***
clarity	35077772957.82	7	3810.2	< 2.2e-16 ***
Residuals	66858452084.92	49981		



Analysis and Modeling

When scoring the model on the new diamonds data source, the scatterplot displays how datapoints are plotted along a tighter line of best fit.

Some of the prices are predicted to be negative since additional predictors would be required to perfect the model.

Recommendation

I would recommend a bid of approx. $8,230,695.69.
I arrived at this number by using the regression model equation provided to predict the price for the set of 3,000 diamonds. Since “the company generally purchases diamonds from distributors at 70% of that price”, I summed up the predicted prices of all set members ($11,758,136.6996) and then multiplied it by 70% to arrive at the final predicted bid mentioned above.