

PREDICTING PRICES OF LOOSE DIAMONDS

LESLIE CARDONE
APRIL 16, 2021



PHOTO BY LEAH KELLEY FROM PEXELS

GOALS

To explore the relationship between the physical properties of loose diamonds and diamond prices

Can we build a highly predictive linear regression model?

Apply the model to loose diamond data obtained from a different source

Can this model predict consistently across a different dataset?



DATA



BLUENILE.COM

RETAILER



52,372 SAMPLES
\$500 - \$13,000
ROUND
GIA GRADED
11 FEATURES

DATA



BRILLIANCE.COM

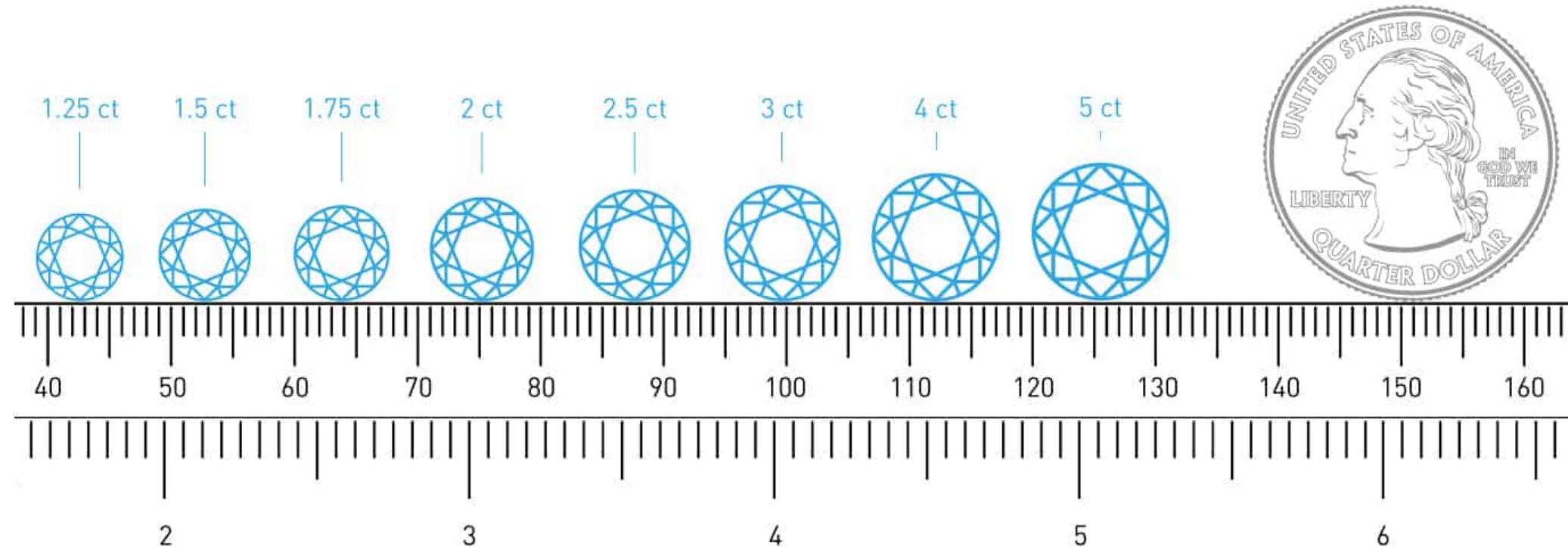
WHOLESALER



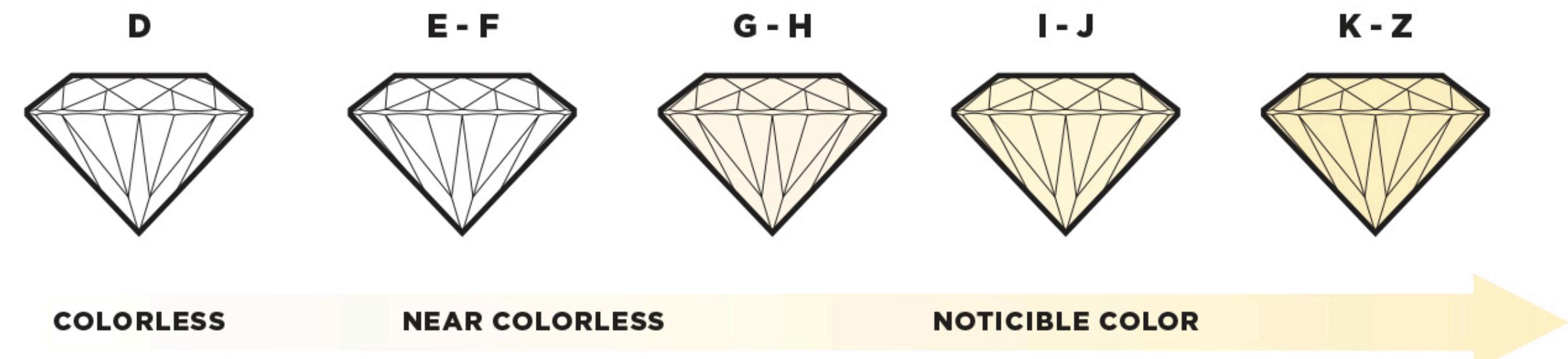
21,782 SAMPLES
\$500 - \$13,000
ROUND
GIA GRADED
11 FEATURES

DATA

CARAT WEIGHT



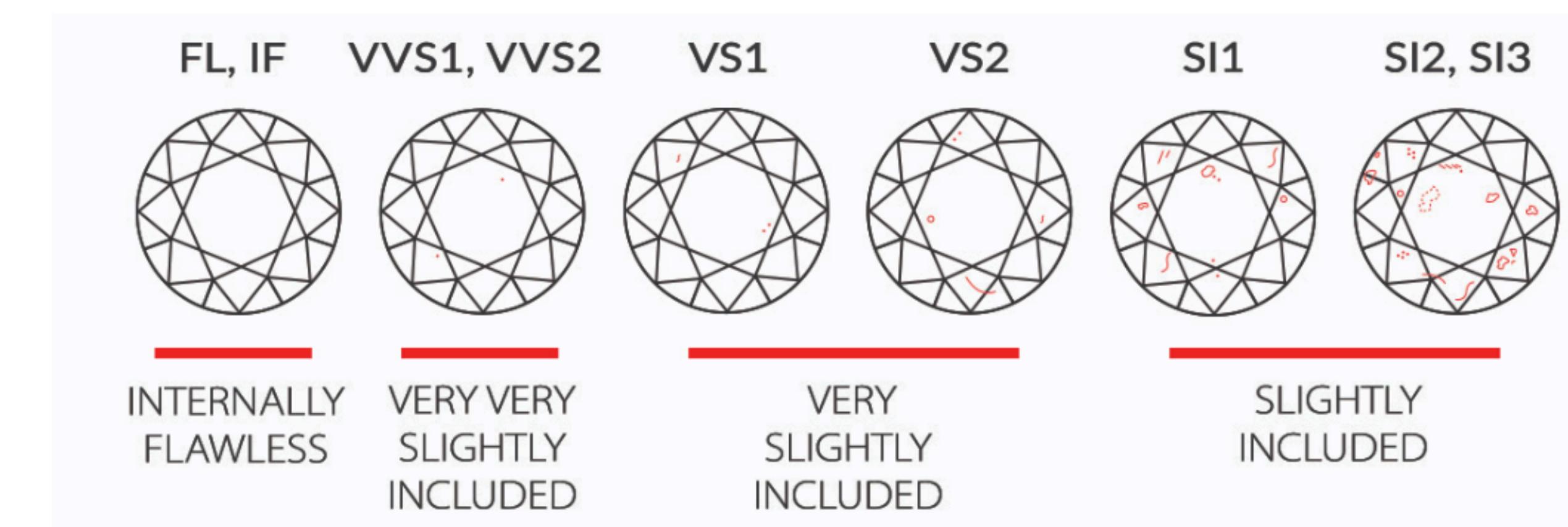
COLOR



CUT



CLARITY



HYPOTHESIS



BLUENILE.COM

RETAILER



BRILLIANCE.COM

WHOLESALE

“
~~In Store Price: \$3,452~~
Our Price: \$2,301

“
~~In Store Price: \$12,000~~
Our Price: \$8,000

The Blue Nile model will consistently 'over-guess' the price for the Brilliance diamonds

METHODS

LINEAR REGRESSION WITH MULTIPLE FEATURES FOR BASELINE

**FIT TO BLUE NILE TRAINING DATA
APPLY TO BLUE NILE VALIDATION DATA
KEEP TEST DATA TO THE SIDE**

TRAINING R-SQUARED: 0.8777...

VALIDATION R-SQUARED: 0.8776...

TRAINING RMAE: \$957.377...

VALIDATION RMAE: \$956.937...

METHODS

FINAL MODEL

CUSTOM FEATURE INTERACTIONS

SKLEARN.PREPROCESSING POLYNOMIALFEATURES

SKLEARN.PREPROCESSING STANDARDSCALER

LASSO REGRESSION

FIT ON BLUE NILE TRAIN + VALIDATION DATA

APPLY MODEL TO BLUE NILE TEST DATA

APPLY MODEL TO BRILLIANCE DATA

FINDINGS

BLUE NILE VALIDATION R-SQUARED: 0.8776...

BLUE NILE VALIDATION RMAE: \$956.937...

BLUE NILE TEST R2: 0.956...

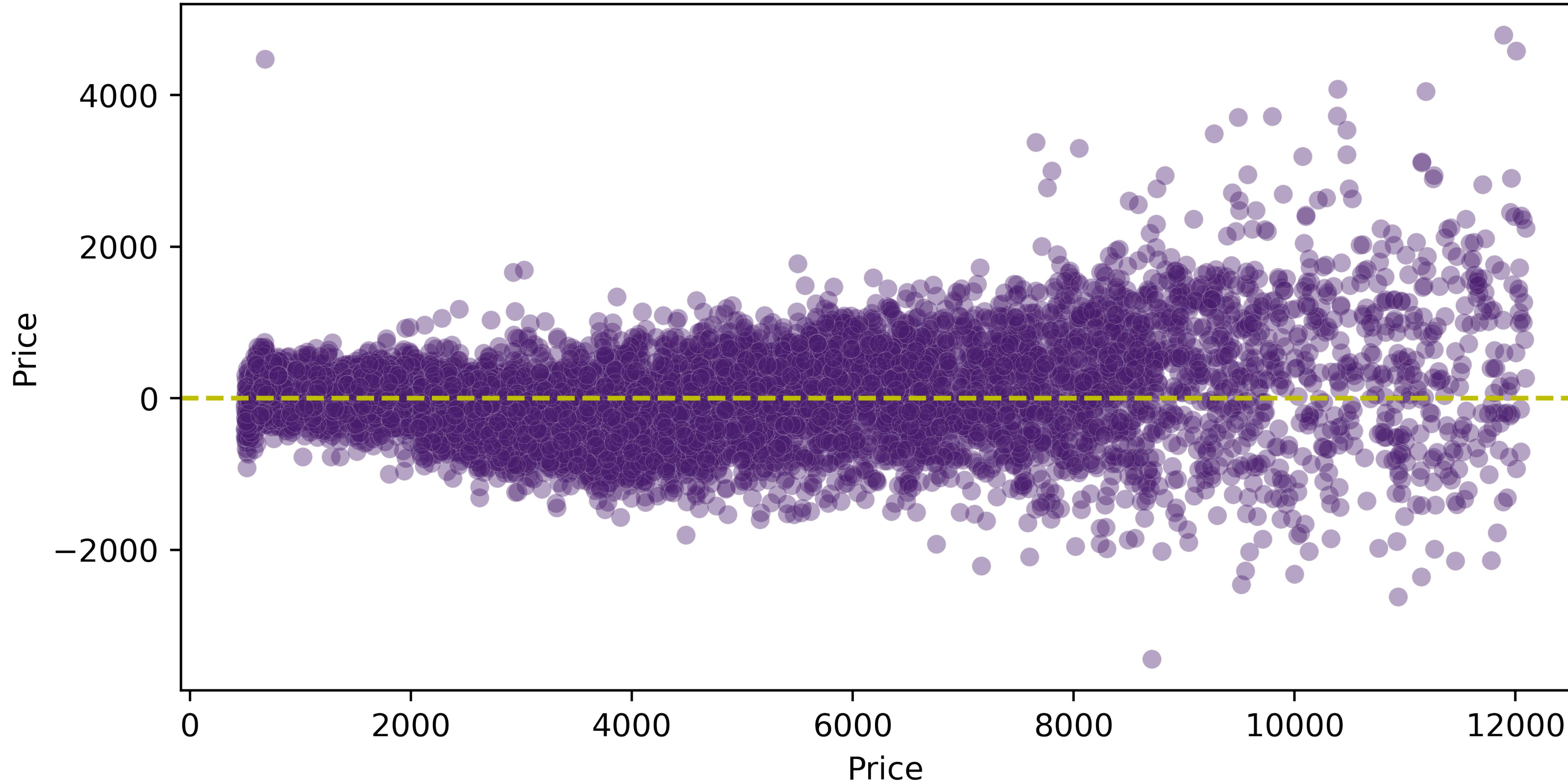
BLUE NILE TEST RMSE: \$564.249...



FINDINGS



BlueNile.com Price Residuals



FINDINGS

BLUE NILE VALIDATION R-SQUARED: 0.8776...

BLUE NILE VALIDATION RMAE: \$956.937...

BRILLIANCE R2: 0.8711...

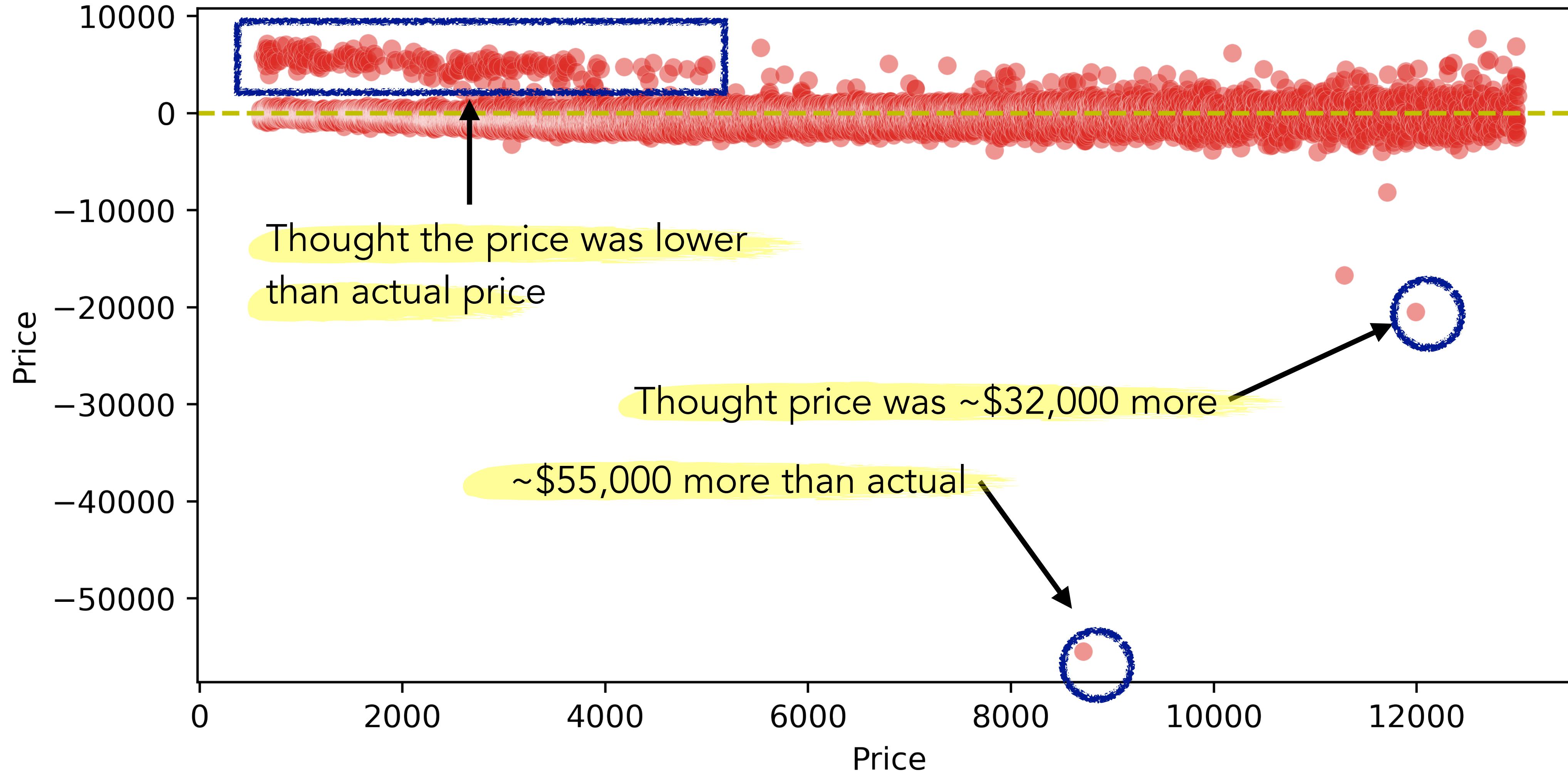
BRILLIANCE RMSE: \$1,003.682...



FINDINGS



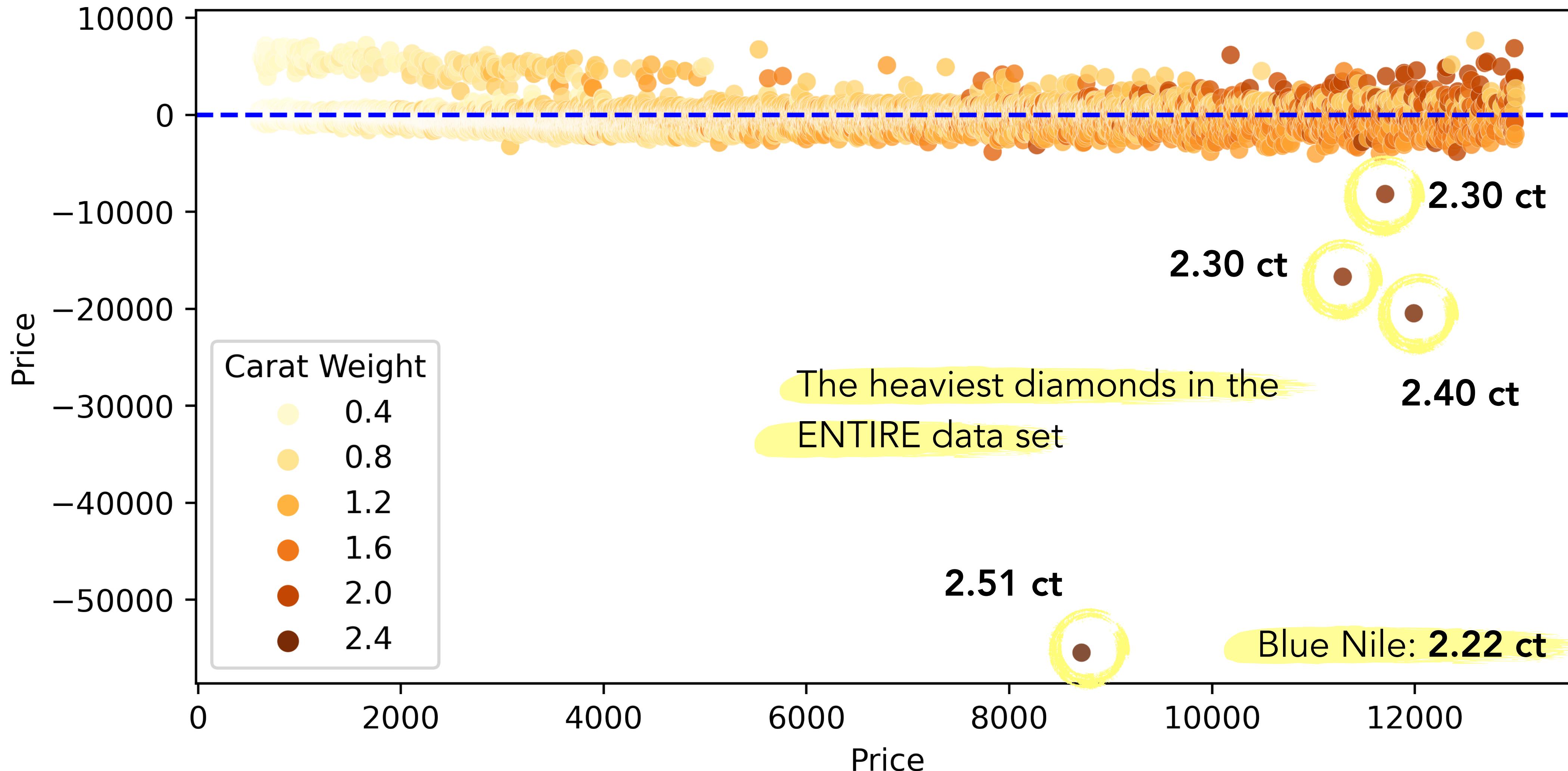
Brilliance.com Price Residuals



FINDINGS



Brilliance.com Price Residuals (Against Carat Weight)



CONCLUSION

Yes, we can create a highly predictive linear regression model for loose diamonds.

Yes, our model can predict fairly consistently across a completely different data set.

No, Brilliance diamonds are not a better deal



FURTHER STUDIES...

Collect more data across a wider range of prices for Blue Nile and Brilliance

Use model on additional diamond distributors to continue to test hypothesis

Try out other models like XGBoost

