Diamond Pricing Analysis and Modeling

By Trinity Gahagen

So you want to get into the diamond business?

- It’s important to understand the different facets of diamond grading
- Namely, the *four “C”s*:
  - Cut
  - Clarity
  - Carat
  - Color
- Cut grades the quality of how the diamond was cut
- Clarity grades how flawless the diamond is
- Carat is the weight of the diamond
- Color grades how *color-less* the diamond is (the less, the better!)

Consider adding a predictive model to your website!

Utilize the four “C”s in a way that is meaningful for your business!

Every major jeweler with a website has a diamond pricing guide, so in order to stay competitive, consider adding a predictive model so that customers can get a sense of which diamonds are within their budget range!

How did I make one?

Testing Results

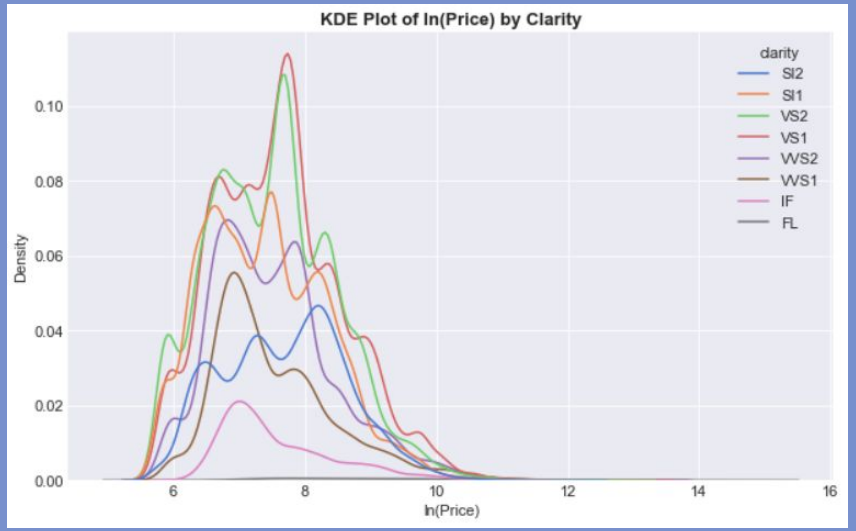
Before beginning the modeling process, several hypothesis tests were performed on diamonds in a dataset to get a sense of what information might be useful when creating the model.

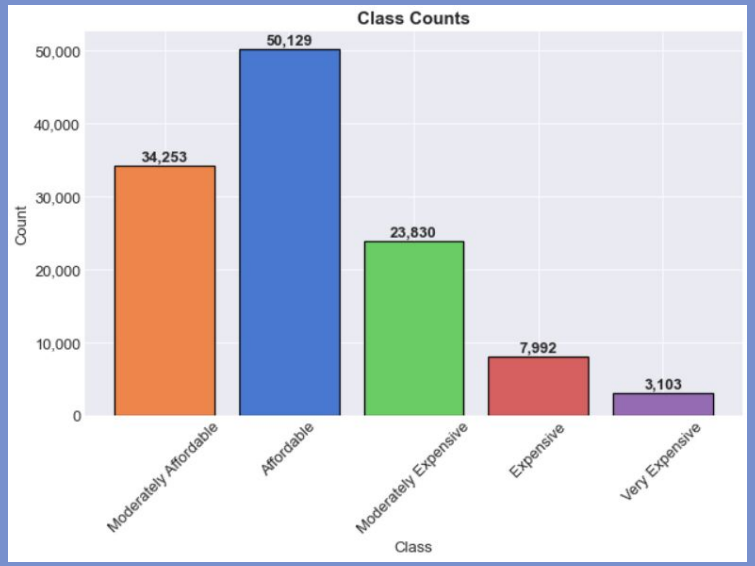
In each categorical variable tested:

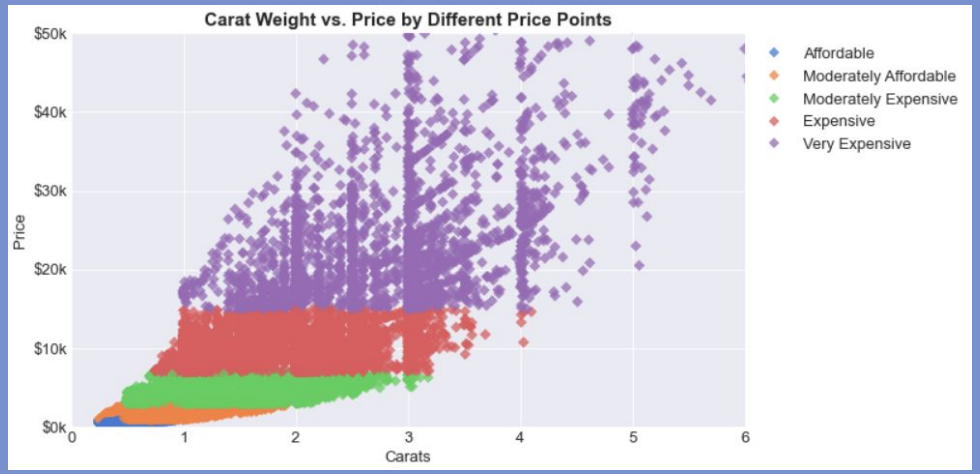
- Shape
- Color
- Cut
- Clarity
- Which lab the diamond was graded at
- Whether the diamond was natural or lab-created

...

All of the tests indicated there was some difference in price between the different grading scales/lab reports EXCEPT for the difference between natural and lab-created diamonds. This means that most features were worth further scrutiny to be considered in the final model.







Target: Price Points

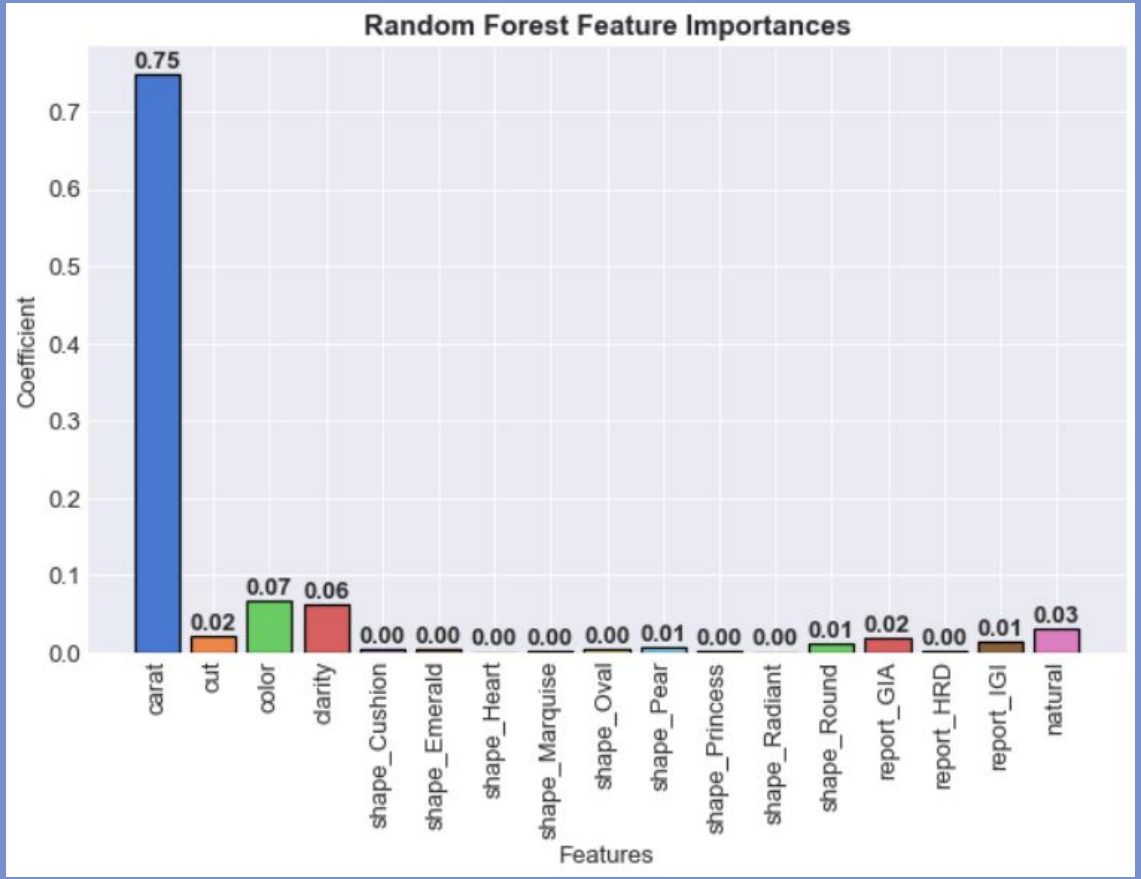
The model I created was a Random Forest Classifier, which assigned price-point labels to diamonds based on the different attributes of the diamond. I engineered the target variable to train/test the model on by assigning different categories of affordability based on price:

- Affordable
- Moderately Affordable
- Moderately Expensive
- Expensive
- Very Expensive

(When this is deployed on your website, the customers can set this range and the model can calculate the thresholds from there)

The Model: Feature Importance

The shape of the diamond had the least impact on the outcome, and the carat weight had the most. I chose to use only the features that had an importance higher than 0.01.

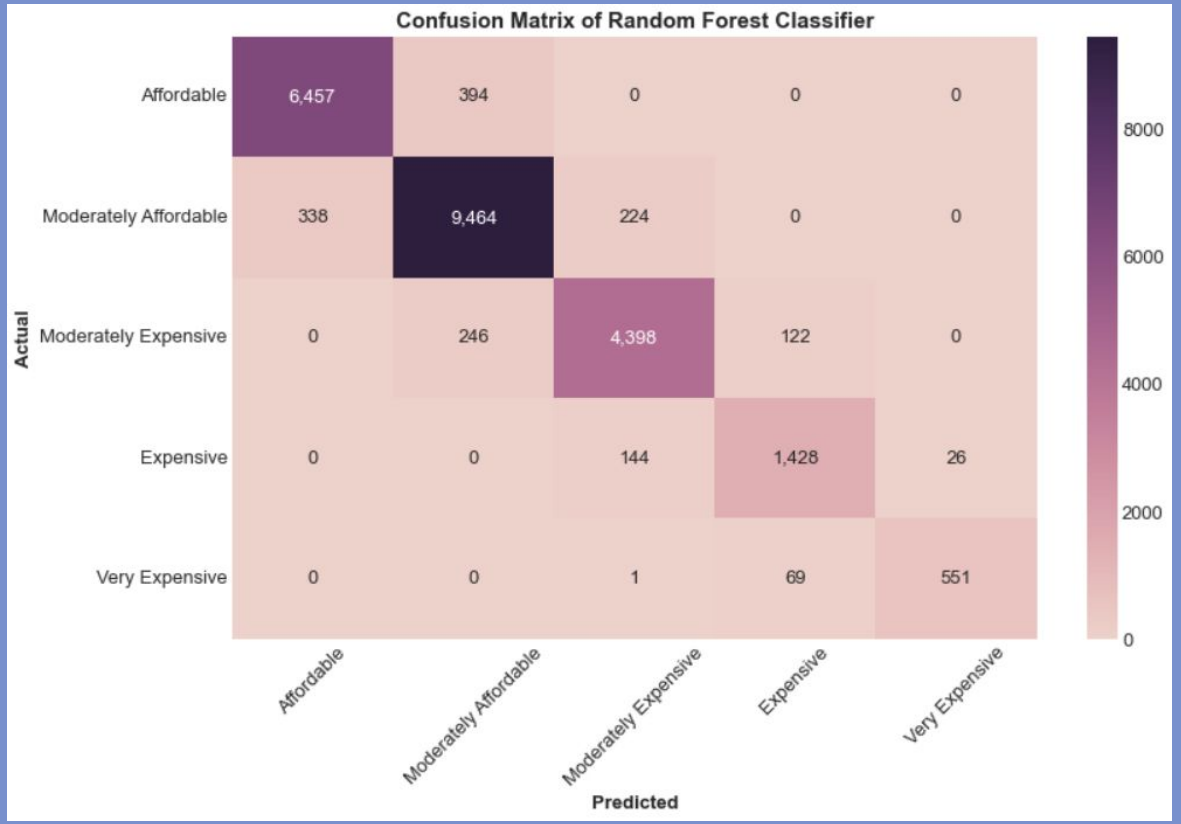


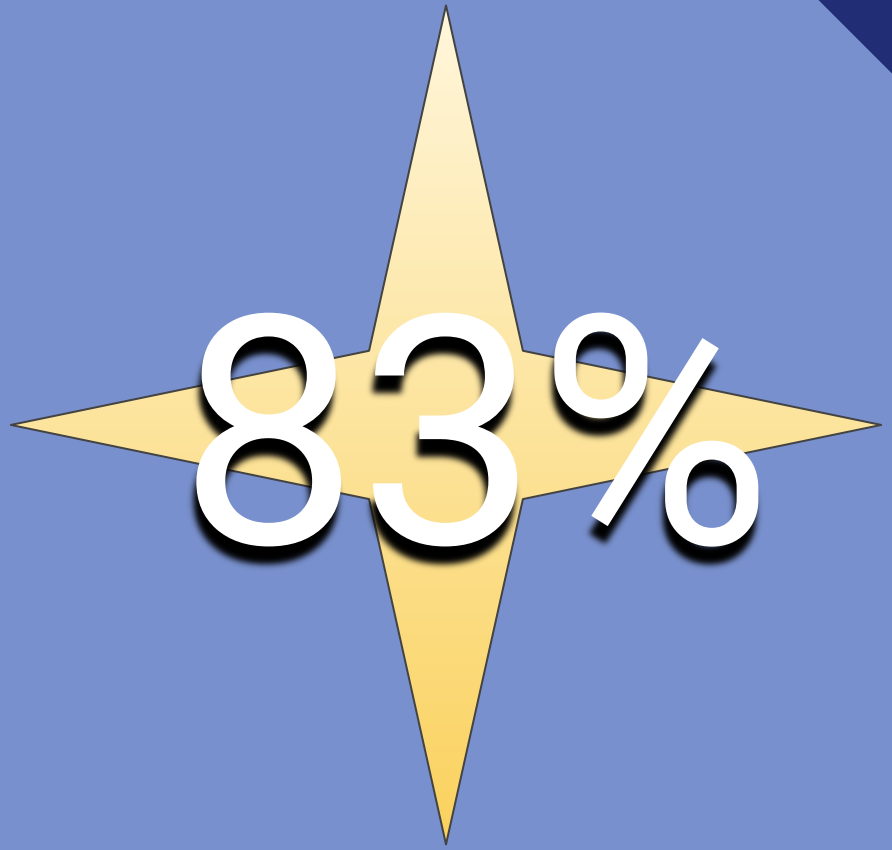
How did the model do?

The *confusion matrix* on the right shows the price points that the model predicted the actual label of the price point.

The model had some trouble predicting points that were very close to the thresholds between each category, but it did an overall decent job at predicting the correct labels.

Overall, the model had an average accuracy score of…





While there is room for improvement, the potential for this model to help customers make their shopping experience better is worth looking into!

With a model like this…

Your customers can enter their budget range on your website and be shown different diamonds that are categorized by affordability. Then, they can easily sort through which ones they are willing to purchase based on much simpler criterion, rather than just manually sorting through all the diamond choices themselves!

Thank you!

External Sources

- Blog at worthy.com:

  https://blog.worthy.com/knowledge-center/diamonds/estimate-diamonds-value/

- Diamonds.pro:

  https://www.diamonds.pro/education/diamond-prices/