Analysis of Blue Diamond Qualities on Price

By Seth Bitney, Sarah Hall, Nathan Koh, and Hannah Richardson

1. Summary

A sample set of over 1,200 diamonds available for sale by Blue Nile was analyzed to determine the role that the 4Cs (cut, color, clarity, and carat) played on diamond price. While it is generally believed (and asserted by Blue Nile) that a diamond's carat is the biggest factor in the price, the company states on their website that the other "C"s can also play a significant role. Each diamond in the sample set had a list price, along with its cut, color, clarity, and carat. Each factor of the 4Cs was analyzed individually, compared to price, as well as to the other factors, in order to better understand how each factor contributed to a diamond's final price. Claims on the Blue Nile website regarding diamond prices were also assessed to determine if their analysis was supported by the pricing of the products within the dataset.

Surprisingly, many of Blue Nile's claims are not reflected in the findings from the sample products. While it was true that carat had a large impact on diamond price, the Blue Nile website information regarding the impact cut, color, and clarity have on price was not corroborated in the data.

Carat, the weight of diamond, was found to have an impact on the price. As a diamond's weight increases, its rarity and price increases rapidly, so carat weight can even be said to have an exponential impact on price. As a result, the website statement that carat was the most important price factor was accurate.

Cut, color, and clarity had less consistent impact on price. Cut, the measure of how well-proportioned a diamond is, showed an inverse trend to Blue Nile’s claims. Despite many examples in the dataset of high-priced diamonds with the highest rated cut, data analysis revealed that the highest average prices were for the lowest rated cut. In other words, as cut quality increased, average price decreased, contradicting some of the website claims. Blue Nile also asserted that a diamond's color affected the price as well. Since colorless diamonds are both rarer and considered higher quality, they state that the most colorless diamonds are the most expensive. However, analysis found that cost did not consistently rise along with the rarity of the color. Additionally, colorless diamonds were not rarer than the other color grades. Finally, Blue Nile also claimed that clarity, the measure of a diamond's imperfections, impacted price, with the price increasing as a diamond's clarity improved. However, the data did not fully support this claim either.

1. Definitions

* **Clarity**: The clarity of a diamond refers to assessing the number of imperfections within a diamond to determine its value against the Diamond Clarity Scale. The scale is defined below:
  + ***Flawless (FL)***
  + ***Internally Flawless (IF)***
  + ***VVS1, VVS2, Very, Very Slightly Included (VVS)***
  + ***VS1, VS2 Very Slightly Included (VS)***
  + ***SI1, SI2 Slightly Included (SI)***
  + ***I1, I2, I3 Included***
* **Color**: Refers to the colorlessness of a diamond. The less color a diamond has, the higher its color grade and value are.
* **Cut**: The cut of a diamond measures how well-proportioned a diamond’s dimensions are, including its balance and brilliance.
* **Carat Weight (Carat)***:* The carat measures a diamond’s weight, not its size. Blue Nile claims “one weight is not necessarily better than another".
* **Price per Carat (PPC):** *Value created for this analysis.* Individual diamond’s price divided by its carat weight.

1. 4C Analysis and Blue Nile Claims

**Clarity Analysis:**

The following Blue Nile claims are based off the assumption “all the other diamond gradings are the same (source).” When looking into the data provided, we cannot prove that this is the case so it will affect the price of some of our observations.

Blue Nile makes the following claims of clarity:

1. The higher the diamond is on the grading scale the higher the value, or price, of the diamond. The grading scale is as follows, in order of lowest quality to highest; SI, VS, VVS, IF, and FL.
2. SI and VS diamonds are the best value because they are less expensive while still having high quality.
3. Color has an impact on clarity because the color changes how easily a “characteristic is seen.”

From the scatter plot (Fig. 1) we can see that despite the differences in the spread of data for all clarity, rankings() are most commonly occurring around the same value, under $50,000 regardless of the clarity classification. At this point we are unable to refute or confirm any claims, so it will be useful to look at the averages of prices to determine how clarity has an effect on price.

From the bar chart (Fig. 2) we can see that the average price of diamonds for each clarity grade should go from lowest to highest quality. Based on the graph we can see that it appears that assumption one is partly met since FL have the most expensive price on average. But, from this visualization it is unclear if VS or VVS is more expensive on average. To look deeper into this assumption a new variable is created that will group classifications that should fall under a similar price range. For example, VS1 and VS2 become the clarity group VS, as seen in Fig. 3.

Based on the grouped bar graph (Fig. 3) the grading scale is as follows, in order of lowest quality to highest; SI, VS, VVS, IF, and FL. The values of the average price of each clarity scale are summarized in figure 4.

Looking at clarity claim 1, we can see that quality does appear to have some impact on price since the FL price average is much higher than all other groupings and it is the highest clarity quality. But we would expect the graph to increase from left to right if the entire assumption is true. We can see that even though IF is the 2nd highest quality it has the 2nd lowest price.

This visualization (Fig. 3) also refutes claim 2, VS and SI are the best value. Best value is classified as the cheaper price for a higher quality diamond. While VS diamonds are significantly lower price than FL, they are on average more expensive than IF diamonds. This would imply that IF diamonds would be the best value because they have a low price on average and are the 2nd highest quality. SI does have the cheapest price point but is only a 43% price difference (calculated based off observations in Fig.4) from IF and it is significantly lower quality.

To investigate claim 3, to determine if color has an impact on clarity, we will compare the different colors that fall under each clarity classification. When visually comparing diamond color by clarity (Fig. 5) we can see that FL only has colors F and D present. This could indicate that these two colors result in a higher clarity but since there are so few diamond observations of FL we cannot come to this conclusion. For all other clarity classifications, it appears that there is an even count of colors that make up the total number of diamonds. Therefore, we can refute the claim that color has an impact on clarity.

**Color Analysis:**

According to Blue Nile, color is the second most important variable out of the 4 C’s when affecting a diamond’s price. In fact, they use colorless to determine a diamond’s color grade and value. As noted in Blue Nile’s education page for Color, the grading hierarchy starts with K color diamonds which are considered to have faint color in them (the budget option) and ends at D color diamonds (the most expensive option) which are claimed to be rare, colorless, and the highest quality of diamond. The full range of this scale is shown below arranged by price in a descending order:

*Colorless diamonds: Rarest and highest quality with a pure icy look.*

* D, E, F Color Diamonds

*Near-colorless diamonds: No discernible color; great value for the quality.*

* G, H, I, J Color Diamonds

*Faint color diamonds: Budget-friendly pick; pairs well with yellow gold.*

* K Color Diamonds

Note: for the purposes of this project, the Blue Nile dataset that was used did not contain any K color diamonds. As such, we will consider J color diamonds to be the budget-friendly option and interpret our results similarly. The claims we want to either prove or disprove are the following:

1. Diamond prices either increase or decrease in alphabetical order. According to Blue Nile, a diamond with a G color grade is less expensive than a diamond with a D color grade
2. D, E, and F color diamonds are rarer than the rest. Blue Nile claims that these color grades are the rarest form of diamonds
3. Lastly, diamond’s price is affected by color

To prove claim 1, we created a bar chart to compare the total number of diamonds across each color grade (Fig. 1). By viewing the chart on its own, we see that the prices for each look to be steadily increasing from left to right until we look at color grade G. At this point, we do not know why the total price of G colored diamonds is lesser than H. So, we will have to perform additional research to determine the cause.

We can start by analyzing the averages across price, total, carat, and the total number of diamonds across each color grade. This information will be used to plot a new bar chart that displays the average price of each diamond color grade.

The bar chart shows the average price of diamonds in each color grade from the lowest to highest price. When looking back at claim 1, we see that this chart does not support the claim that prices increase or decrease in alphabetical order. We can also see that F and H colored diamonds are affected in the price hierarchy. These can be for the following reasons:

* The total number of diamonds in each color grade are not equivalent. Based on the number of diamonds, it will affect the total, and average, price of each diamond
* Depending on the shape of the diamond, prices tend to fluctuate despite their color grade. According to Blue Nile, the best shapes are Radiant, Cushion, and Princess, which are known to mask the color best. This implies that the diamond shape can affect price despite their color grade
* Lastly, the type of diamond cut will affect diamond price. Blue Nile claims that Astor and Ideal cuts are the most expensive due to these cuts maximizing interaction with light

Therefore, based on these observations, we cannot confidently support the claim that the price of Blue Nile’s diamonds either increase or decrease in alphabetical order.

The second claim we want to prove is the claim that D, E, and F color diamonds are rare. We can look at the proportion table (Fig. 2) and observe the total number of diamonds across each color grade. We see that there are 227, 223, and 181 diamonds respectively. Interestingly, Color J has the least number of total diamonds with 90. This means that J colored diamonds would be rarer than what the claim suggests.

Lastly, the third claim we want to discuss is if color affects a diamond’s price. We can start by observing the following two boxplots (one with outliers and one without). Note that due to extreme outliers across all color grades, it made it difficult to see the boxplots themselves.

From the boxplots, we see that there are significant differences across each boxplot’s interquartile range due to a larger data spread. However, the median line tends to be similar across all of the color grades. This could mean that color may not be as important of a variable as Blue Nile would suggest.

Therefore, analyzing the other variables is encouraged to determine which property affects a diamond’s price.

**Cut Analysis:**

Blue Nile suggests that cut is the most important of the four Cs and recommends prioritizing cut above all other factors. According to their website: "When people talk about someone’s engagement ring, everyone focuses on carat. In reality, it’s the cut of the diamond that can be the biggest factor in the price tag." Blue Nile categorizes diamonds into four cut groups: Astor Ideal, Ideal, Very Good, and Good, with Astor Ideal as being considered the best.

We explored cut data to determine the spread of the values, noting that Astor Ideal has the fewest outliers and the tightest series of datapoints. Interestingly, these are on the lower end of the price spectrum. We explored average price data to better understand how cut impacts price.

The data does not support Blue Nile’s claim of cut being the biggest predictor of price. Based on our analysis, while Blue Nile’s highest grade of cut ("Astor Ideal") has the highest median price ($2,854.00), it has the lowest average price ($5,851.55). The cut data shows an inverse correlation between average price and cut, with the highest average price within Blue Nile's lowest cut ("Good"). While "Astor Ideal" has a small sample size (n=20), the second highest grade of cut ("Ideal") has the largest sample size (n=739) but has the lowest median price ($1354.00) and the second lowest average price ($6,488.93). This suggests that diamond cut does not drive price as Blue Nile claims.

**Carat Analysis:**

Carat refers to the weight of a diamond, rather than the size (as is often assumed). Carats are measured either in milligrams or points. One carat is equal to 200 mg, or 100 points.

Looking at the dataset of Blue Nile diamonds, we can see that over 70% of the diamonds have a carat of less than one. The average carat weight was, in fact, only 0.813, with the median even lower at .52. Consequently, as carat size increased, the number of diamonds decreased. Only 13 diamonds in the dataset had a carat of 5 or more.

Blue Nile really only makes one definitive claim about carat weight- that carat has the largest impact on the price of a diamond. Plotting the carat size against price shows us visually that carat size is seemingly correlated with price.

Blue Nile makes no other clear claims about carat weight, but touches on two other topics. The first is that "fancy shapes cost less per carat". Since we lack the shape information to make this determination, we have no established method to test this.

The second claim is about best value for carat price. Blue Nile advises to buy "slightly below the whole and half-carat marks" in order to "save a considerable amount of money". This is difficult to definitively state, but by plotting the price per carat (PPC) by carat weight, it does appear that the price per carat jumps up at the .0 and .5 marks along the x-axis, with small dips in between.

This chart also further supports the claim that price is related to carat weight, since the heavier diamonds have higher PPCs.

Further analysis of carat weight shows that the average PPC of diamonds with an Astor Ideal cut (considered the best cut Blue Nile offers) is much higher than the other cuts. This would make sense, as a larger or more expensive diamond would be used with the highest-grade cut. Interestingly, Ideal cut has the lowest average PPC, contradicting the website’s claims.

1. **Regression of Price Against Carat**

We want to explore how the carat weight of a diamond may be related to price using simple linear regression. The linear regression model is based on assumptions that must be met to verify that any observations made on the variability of our model are reliable. The assumptions in terms of our analysis are as follows:

1. The errors for each fixed value of carat weight have a mean value equal to zero.
2. The errors for each fixed value of carat weight have a constant variance.
3. The error values are independent.
4. The error values for each fixed value of carat weight are normally distributed.

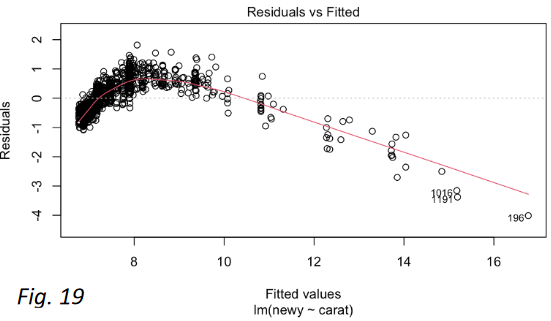
Claims 1 and 2 are the most important so we will only address those two claims. To assess if assumptions 1 and 2 are met by looking into the scatter plot of price against the carat weight (As seen in the carat section on fig. 14).

From that graph, we can see that neither assumption 1 or 2 is met. Assumption 1 is disproved because the data points are not evenly scattered on both sides of the regression line as we move from left to right along the graph. Assumption 2 is not met because the vertical spread of the data points appears to be increasing as we move along the graph. To further investigate the assumptions, the residual plot can be assessed (fig. 17).

We have further evidence that neither assumption is met in this graph. The variance increases with the fitted Y, so assumption 1 is not met. Also, the residuals are not evenly scattered across the horizontal axis so assumption 2 is not met. Since neither assumption 1 nor 2 is met we will start by transforming our response variable (price) since doing so affects both assumptions and will stabilize the variance.

Because the residuals are increased when moving along the graph, a log transformation with a lambda value less than one is an option for transforming price. To confirm that we can transform on the price we must first graph a box-cox plot to verify that a lambda value of 1 does not exist inside the confidence interval for the graph of log-likelihood function against lambda.

From the box-cox plot (Fig. 18) we see that lambda equal to 1 does not fall under our confidence interval. Therefore, we can perform a log transformation on price. To transform the response variable our new y value, y\* will equal log(y). So, our new value for price is price\* = log(price). To verify that our transformation improved our model we will reduce price\* against the carat weight.



A new residual plot is created (Fig.19) to check the first two assumptions. From the residual plot we can see that the variance of the residuals improved, in other words the variance is more constant when moving from left to right on the graph. Based on the residual plot the residuals are still not evenly scattered across the horizontal axis, indicating that there is a non-linear relationship between the variables. Assumption 1 is not met so, we will need to transform carat, the x variable, next. To determine how to transform the carat weight value, we will need to create a scatter plot of our adjusted price value, price\*, against carat weight.

From the scatter plot (fig. 20) of our adjusted price against carat weight, we can see that the scatter plot appears to follow a log distribution so a log(x) transformation will be applied to the carat weight. In other words, our new value for carat weight will be carat\* = log(carat), where carat represents the carat weight. After regressing price\* against carat\* and graphing the scatter plot (fig. VIX) we can see that it appears that our model improved and assumptions 1 and 2 are now met.

To further verify our thoughts that assumption 1 and 2 are met we will investigate the residual plots of our adjusted model with price\* and carat\*.

From the residuals graph in the top left of Fig. 22, we can see that both assumption 1 and 2 are met, so our transformation worked. Now the residuals are evenly scattered across the horizontal axis with no pattern and the vertical spread is constant as we move from left to right on the graph. From the scale-location plot (bottom left of Fig. 22) we can see that the vertical variance appears to be consistent for all fitted values of carat weight so that further verifies that assumption 2 is met.

For our linear regression to determine if there is a linear relationship for price and carat weight from our adjusted model we will perform and ANOVA F Test. This test determines if there is a linear association between carat weight and price by testing whether the slope of our simple linear regression equation is equal to 0. If the slope is zero then the predictor (carat weight), has no effect on the price. In that case there would be no linear relationship.

The null hypothesis we will be testing using the ANOVA F-test is if our slope is equal to zero. Our slope is defined as the change in price on average when the carat weight increases by one unit. The alternative hypothesis is that the slope is not equal to zero.

Walking through calculations for the regression. The results of the final regression are summarized in Fig. 23.

Additional details can be found in the ANOVA table for the regression (Fig. 24)

The value for our test statistic for the ANOVA F test is F=MSR/ MSres. From our ANOVA table (fig. 24) we can see that the anova F statistic is 25,535 and the corresponding p-value is 2.2e-16. Since the p-value is very small, we reject the null hypothesis. In other words, we can support the claim that there is a linear relationship between price and carat.

Proving this claim can also be accomplished using a critical value test. Using R, we can determine that the critical value is 3.849. The F-statistic is far larger than the critical value, so we can reject the null hypothesis using this test as well. We can claim that there is a linear association between carat and price.

Looking at our summary results, we can see that R2 is .9547. In context, this means that 95.47% of the variance in price in our model can be attributed to carat weight.

Our final linear formula is as follows: **y\*=1.94x\* + 8.52,** where we let y\* be *log(y)* and x\* be *log(x)*. Based on our regression analysis we can conclude that there is a linear relationship between carat weight and the price of a diamond.

1. **Conclusions and Limitations**

Many of Blue Nile's claims are not reflected in the findings from the sample data. Although carat had a predictable impact on diamond price (carat and price have a linear relationship) the Blue Nile website information regarding the impact cut, color, and clarity have on price was not corroborated in the data.

Analysis was done using Simple Linear Regression (SLR) methods which cannot take into account interactions between predictive variables as accurately as Multiple Linear Regression (MLR). Additionally, varying levels of objectivity exist within the 4Cs in diamond pricing. For example, carat is the most objective measure based exclusively on carat weight, whereas cut is the least objective: the ideal cut of each diamond will vary slightly, and attributes such as amount of sparkle and fire are difficult to objectively measure. As a result, some unavoidable nuance exists within the data.

Finally, as the data comprised diamonds for sale (as opposed to diamonds already sold), it does not provide a perfect representation of what consumers may be willing to pay based on the 4Cs. As such, diamonds within the dataset may be underpriced or overpriced for the market, impacting overall analysis. Further analysis could also utilize MLR to better understand interaction between predictive variables on diamond pricing, highlight collinearity, and provide even more detailed analysis to better serve Blue Nile and its customers.

Sources

All data and definitions were pulled from Blue Nile’s website.

*4Cs of Diamonds & Diamond Buying Guide*. Blue Nile. (2024). <https://www.bluenile.com/education/diamonds>