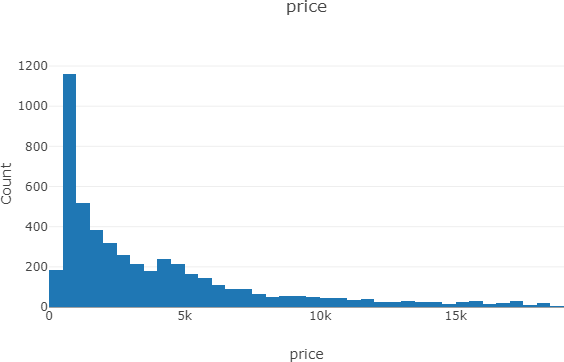
*Evaluating Properties in Relation to Diamond Price*

INTRODUCTION

Diamonds have gained a large popularity in today’s society and are considered to be highly luxurious gemstones. Society’s specificity associated with the qualities of diamonds have become increasingly particular. This makes the search for suitable diamonds difficult and impacts the price of the diamond. The goal of this project is to better understand the relationship between different qualities of a diamond and the diamond’s price. This work may be educational for those in the market for a diamond of some sort and may help provide an expectation for monetary association to diamonds based on some specific factor.

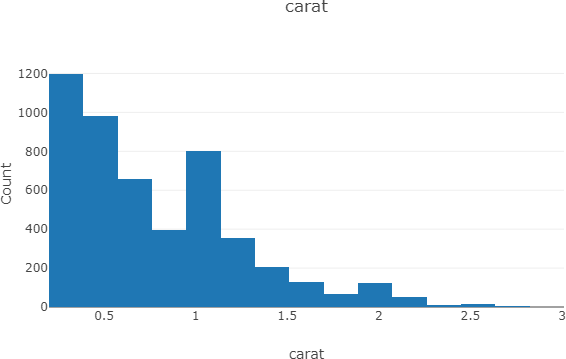
DATA

Our data comes from a sample of 5000 diamonds, with different prices and qualities recorded, the source of the data however, is unknown. The focus of this project will center around the quantitative variable price which measures the final cost of the diamonds(in $) included in the dataset. This dataset has an average price of $3859.48 and a standard deviation of $3906.04. The distribution of price in this dataset is unimodal and skewed to the right.



The research questions will also utilize the variables of diamond color, cut, and carat. Diamond color is a categorical variable indicating the quality of the color of the diamond, assuming that pure diamonds(colorless and clear) are of the best quality. In this dataset, color is most often graded a G(1034 of the diamonds were graded a G, making up 20.7% of the 5000 diamonds sampled. Diamond cut is a categorical variable that measures the quality of the diamond’s cut, assuming that a well-shaped diamond free of impurities is of the best quality. In this dataset, diamond cut is most often rated “Ideal”(1930 of the

diamonds were rated “Ideal”, making up 38.6% of the sampled diamonds). Lastly, carat refers to the weight of the diamond(1 carat=200 mg). This dataset has an average weight(in carats) of 0.793 and a standard deviation of 0.469. The distribution of carats is essentially unimodal and skewed to the right.



RESEARCH SCENARIOS

*CUT*

Our first research question is to investigate whether there is a relationship between price and cut of the diamond and if so, characterize the association. If there are varying degrees of diamond cut value, where some are better than others, it seems reasonable to assume that the diamond’s price would reflect this. To start off, in the table below, diamond price is summarized by each rank of cut.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Fair | Good | Very Good | Premium | Ideal |
| Mean Price($) | $3737.58 | $3683.55 | $3874.55 | $4583.57 | $3402.21 |
| St Dev Price($) | $3144.08 | $3420.19 | $3837.02 | $4287.51 | $3759.07 |

From the table, we can see slight deviation amongst the average prices of differently ranked cuts. The premium cut has the highest average price, $4583.57, while an ideal cut has the lowest average price,

$3402.21. To further investigate this, we can look at the boxplot given below.



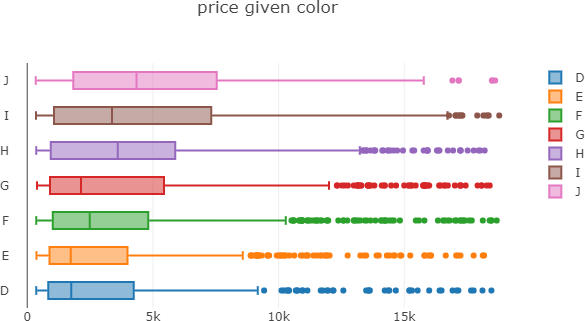
We can see from the plot that the median price for the premium cut is higher than the average prices of the rest of the cuts, and the median price for the ideal cut is lower than the rest of the average prices. This is consistent with the mean prices shown in the above table. But overall, the interquartile ranges demonstrate a lot of overlap and therefore we cannot reasonably assume a relationship between diamond cut and price.

*COLOR*

For our second research question, we are investigating whether there is an association between color and price of a diamond. If there is a relationship between the two, we are investigating how to characterize that. Since there are colors that are characterized as good and bad, an association between color and price seems like a valid thought. As we did with diamond cut, below is a table that summarizes the average price for different color grades of diamonds(D=best, J=worst).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | D | E | F | G | H | I | J |
| Mean Price($) | $3133.24 | $2994.31 | $3745.82 | $3792.00 | $4427.27 | $5033.21 | $5378.54 |
| St. Dev Price($) | $3493.75 | $3182.18 | $3718.21 | $3880.69 | $4034.91 | $4724.33 | $4517.34 |

From this table, we see some variation across the average prices of differently graded colors. Grade E has the lowest average price of $2994.31 while grade J has the highest average price of $5378.54. We can look at the boxplot for the two variables to look further into this.

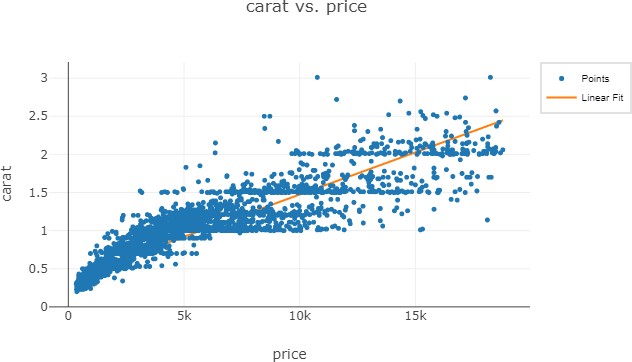


This plot shows the median price of diamonds graded a J is higher than that of the others, similarly, the median price of diamonds graded E is lower than that of the others. These depicted medians are similar to the previously identified mean prices of each color grade. Overall though, there is significant overlap between the different color grades and therefore we cannot establish a correlation between price of a diamond and its graded color.

*CARAT*

Our final research question is whether or not there is a relationship between diamond price and carat and if so, to investigate how to characterize the association. As one of the four key properties of a diamond, carat seems like it might have a relationship to a diamond’s price.

To evaluate the potential relationship between these two variables, a scatterplot has been constructed.



As shown above, we plot carat against diamond price. There seems to be a reasonably linear, positive relationship between the two variables on the graph. Below, the correlation coefficient and linear regression model is noted below. The correlation coefficient of 0.922 suggests a strong positive relationship between carat and price.

|  |  |  |
| --- | --- | --- |
| Variable: | Correlation | N |
| carat vs. price | 0.922 | 5000 |

Regression of carat on price (model id: slm4)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Std. Error | t | p-value |
| Intercept | 0.3658 | 0.0036 | 101.3421 | 0 |
| price | 0.0001 | 0.0000 | 168.4388 | 0 |

DISCUSSION AND CONCLUSION

The analysis of this project focused on the relationship between the price of a diamond with a number of diamond properties from a sample of 5000 diamonds. Specifically, the potential relationships between price and diamond cut, price and diamond color, and price and diamond carat were investigated. There was no strong relationship identified between diamond price and color or diamond price and cut. However, as anticipated, there was a strong relationship between diamond price and carat. Future research should evaluate clarity, the final of the typical diamond properties and its relationship to price, we would expect there is a positive relationship between diamond clarity and price. It may also be important to further evaluate the

source of the sample, in this case, it is unknown but for future analyses ensuring a random, representative sample is imperative to the accuracy of the research. Overall, the analysis is able to assist consumers and sellers alike in understanding what factors are related to diamond prices.