Paul Flanagan

Diamond Data Analysis Project Summary

Tools Used: pandas, seaborn, matplotlib

Questions Asked: 1. What are the standard dimensions for a diamond being sold?

2. What is the distribution of carat sizes typically circulating in the market?

3. How are the tiers of qualities distributed?

4. How rare are the highest tiers of quality in a diamond?

5. How does a diamond's value correlate with each of the main attributes?

Insights Discovered: Average dimensions are 5.73mm x 5.73mm x 3.54mm, with right-skewed distributions for each measurement, meaning that the majority of records fall below these averages, but the most distant outliers are above the averages.

The average carat diamond is around 0.75 carat, with a notable but unexplained spike at around 0.3 carat.

A plurality of diamonds in the set has the highest cut quality, while the highest quality color and clarity are much rarer.

48% of the diamonds in the set have at least one of ideal cut, perfect color, or flawless clarity, 8% have at least two, and 0.05% have all three.

The strongest indicator of a diamond’s value is its carat measurement.

Recommendations: I would recommend using this data to better appraise the value and rarity of diamonds encountered through transactions, and to also consider their own stock and how it compares to the distributions of attributes found within this data set.

Future Work: If I were to continue, I would like to be able to identify the impact of more factors on a diamond’s price beyond its carat measurement, perhaps by normalizing the carat size over a sample first. I could also perhaps try to implement a model capable of predicting or suggesting the value of a diamond based on provided measurements and qualities.