

## GIA Diamond Analysis

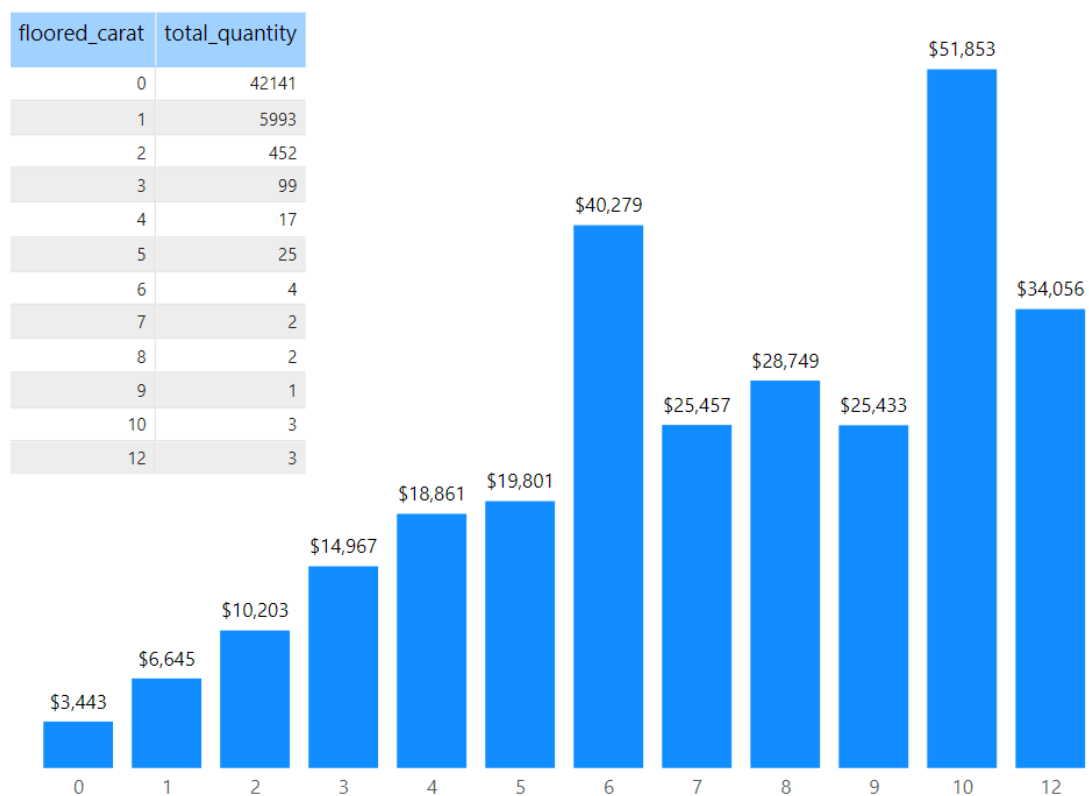
### Abstract

This report statistically analyses the round, natural diamonds from the GIA report, aiming to quantify the effect of Cut, Clarity, Colour and Carat of a diamond and then use the information to help different types of diamond buyer to make informed decision.

**Techniques:** linear regression, Python, SQL, Power BI

**Data Source:** Kaggle

### Rarity



As can be shown from the graph, the rarity of the diamond increase with the weight. There are over 40000 diamonds below 1 carat, while there are only 6000 between 1 and 2 carats. When it goes to 3 or more carat, there are only very few to be found.

This also leads to the value of the diamond increase sharply with the weight. For diamonds below 1 carat, the average price per carat is around \$3,400 dollar, while the price also doubled for diamonds between 1 and 2 carats. Although it is very rare that diamonds weighted above 4 carats, but for diamonds below 4 carats, the average price per carat increases at around \$3000 to \$5000 between each group.

OLS Regression Results						
Dep. Variable:	price_per_carat		R-squared:	0.608		
Model:	OLS		Adj. R-squared:	0.608		
Method:	Least Squares		F-statistic:	7.574e+04		
Date:	Wed, 21 Apr 2021		Prob (F-statistic):	0.00		
Time:	17:48:56		Log-Likelihood:	-4.1899e+05		
No. Observations:	48742		AIC:	8.380e+05		
Df Residuals:	48740		BIC:	8.380e+05		
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	1602.1164	10.380	154.353	0.000	1581.772	1622.460
carat	4058.0351	14.745	275.208	0.000	4029.134	4086.936
Omnibus:	44957.635		Durbin-Watson:	1.718		
Prob(Omnibus):	0.000		Jarque-Bera (JB):	16404936.460		
Skew:	3.665		Prob(JB):	0.00		
Kurtosis:	92.576		Cond. No.	3.43		

From a statistical perspective, if a diamond's weight increase by 1 carat, the value would increase by \$4058.

## Clarity, Colour and Cut

If you have ever talked to a diamond specialist, he would tell you that each of the 4 C's (Cut, Colour, Clarity and Carat) play a role in a diamond's beauty.

In this sector, we are going to analyse the effects of the 4 C statistically.

OLS Regression Results						
Dep. Variable:	price	R-squared:	0.572			
Model:	OLS	Adj. R-squared:	0.571			
Method:	Least Squares	F-statistic:	3612.			
Date:	Wed, 21 Apr 2021	Prob (F-statistic):	0.00			
Time:	18:16:24	Log-Likelihood:	-4.8875e+05			
No. Observations:	48742	AIC:	9.775e+05			
Df Residuals:	48723	BIC:	9.777e+05			
Df Model:	18					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-4451.2257	616.555	-7.220	0.000	-5659.681	-3242.770
color[T.E]	-202.6513	73.955	-2.740	0.006	-347.604	-57.699
color[T.F]	-721.4951	77.548	-9.304	0.000	-873.490	-569.500
color[T.G]	-1266.2327	82.830	-15.287	0.000	-1428.581	-1103.884
color[T.H]	-2368.9942	96.242	-24.615	0.000	-2557.629	-2180.359
color[T.I]	-3407.9489	110.004	-30.980	0.000	-3623.557	-3192.341
color[T.J]	-4262.7010	130.358	-32.700	0.000	-4518.204	-4007.198
clarity[T.IF]	-1397.4242	425.338	-3.285	0.001	-2231.092	-563.757
clarity[T.SI1]	-3611.8321	418.963	-8.621	0.000	-4433.005	-2790.659
clarity[T.SI2]	-5198.7102	422.126	-12.316	0.000	-6026.082	-4371.338
clarity[T.VS1]	-2572.3238	417.369	-6.163	0.000	-3390.373	-1754.275
clarity[T.VS2]	-2816.1405	417.572	-6.744	0.000	-3634.588	-1997.693
clarity[T.VS1]	-1997.2595	418.653	-4.771	0.000	-2817.824	-1176.695
clarity[T.VS2]	-2246.5950	418.041	-5.374	0.000	-3065.961	-1427.229
cut[T.Good]	1007.7478	482.954	2.087	0.037	61.153	1954.343
cut[T.Ideal]	2391.0333	460.484	5.192	0.000	1488.478	3293.589
cut[T.Super Ideal]	1626.0845	455.493	3.570	0.000	733.313	2518.856
cut[T.Very Good]	1473.4551	458.645	3.213	0.001	574.505	2372.405
carat	1.65e+04	65.356	252.530	0.000	1.64e+04	1.66e+04
Omnibus:	158825.839	Durbin-Watson:	0.182			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	43078890106.623			
Skew:	55.153	Prob(JB):	0.00			
Kurtosis:	4607.279	Cond. No.	72.6			

As shown above, we can identify that different level of Colour and Clarity would cause a significant price difference:

- For colour, the average price difference for 1 carat diamond between D (highest ranking) and J (lowest ranking) is \$4263.
- For clarity, the average price difference for 1 carat diamond between FL (highest ranking) and SI2 (lowest ranking) is \$5199.
- However, the effect of cut to diamond's price is not that significant, the price difference between highest and lowest ranking is around \$1000. This does not mean that cut is not important, the details would be discussed in the next section.

### Real Case Scenario

After we discussed the statistical effects on the 4 Cs, we can now try to apply these knowledges into some real case scenarios.

As a general rule in the world of diamond, a high-quality Cut above all else for most buyers—as this greatly impacts a diamond's beauty and brilliance. The second most important aspect to focus on is Colour as a diamond should look white or colourless to the naked eye. In terms of Clarity, any diamonds' Clarity above (include) VS2 are believed to be eye clean.

Let us set a budget of \$20,000 and find out what diamonds could be recommended for different kinds of buyers.

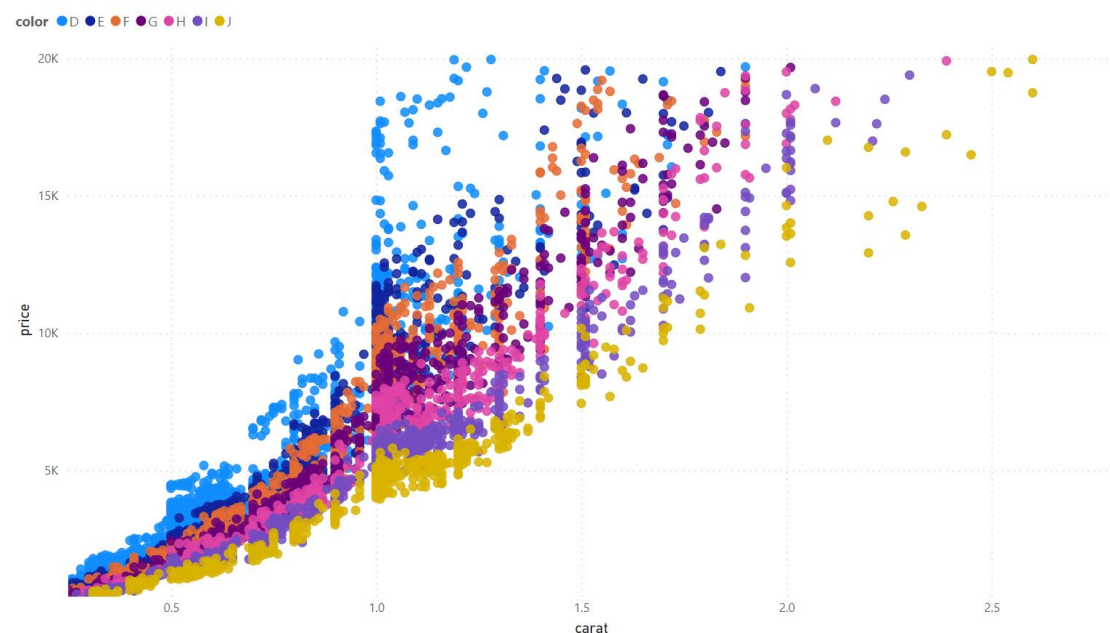
### Scenario A: the buyer is looking for a diamond for the wedding band.

As the diamond is to be decorated on a wedding band, the beauty and brilliance of the diamond play an important role.

In this circumstance, the Cut quality should be high, considering the minor effect of cut on the price but the major effect of beauty and brilliance, it can be recommended to looking for Ideal and Super Ideal only (two 2 levels of cut).

For clarity, diamonds above VS2 could be recommended as they are all eye clean.

The rest is a trade off between colour and carat.



As shown above, we can find that there distribution of the diamonds with different colour is very clear. The higher the carat, the lower level the colour could be bought. So, the buyer needs to have some own comparison and consideration to make the final decision.

### Scenario B: the diamond investor

The above graph could also help to find out the investing opportunities. We can identified that for relatively good colours (D, E and F), the prices increase sharply while the weight of the diamond reach 1 carat. Especially for colour D, the 1 carat diamond varies from 11k to around 18k. This represents that the market of D-colored diamonds is the least

transparent, which represents potential investment opportunities.