

# Diamonds

Linear Regression modelling to predict  
the diamonds prices\*

Francesca Logiacco & Hilal Işık

\*This project is created and shared by the mentioned authors, with each joint author enjoying an equal interest in the undivided whole.



# Data Description

---

- Scrapped from Australian Diamond Importers website on 24th Feb 2022
- Includes diamond features, like shape, carat, colour, cut, clarity, and other features
- The price in AUD (excluding GST)
- Data size: 26.6 Mb; 219704 rows; 27 columns.



# Data Cleaning

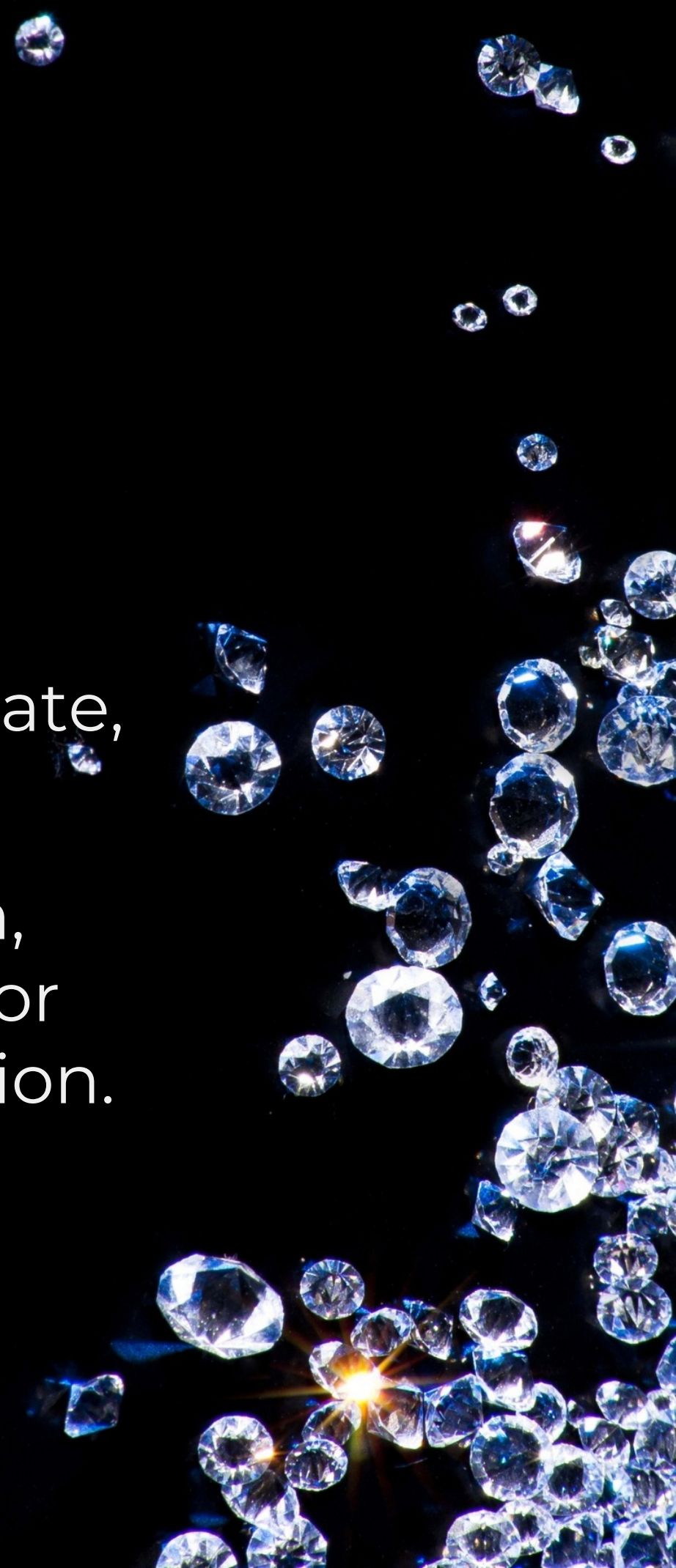
---

## Dropped 10 columns

- 3 columns with no impacting features: diamond id, date, lab.
- 7 columns with more than 70 % null values: eye clean, fancy color dominant, fancy color secondary, fancy color overtone, fancy color intensity, fluor color, culet condition.

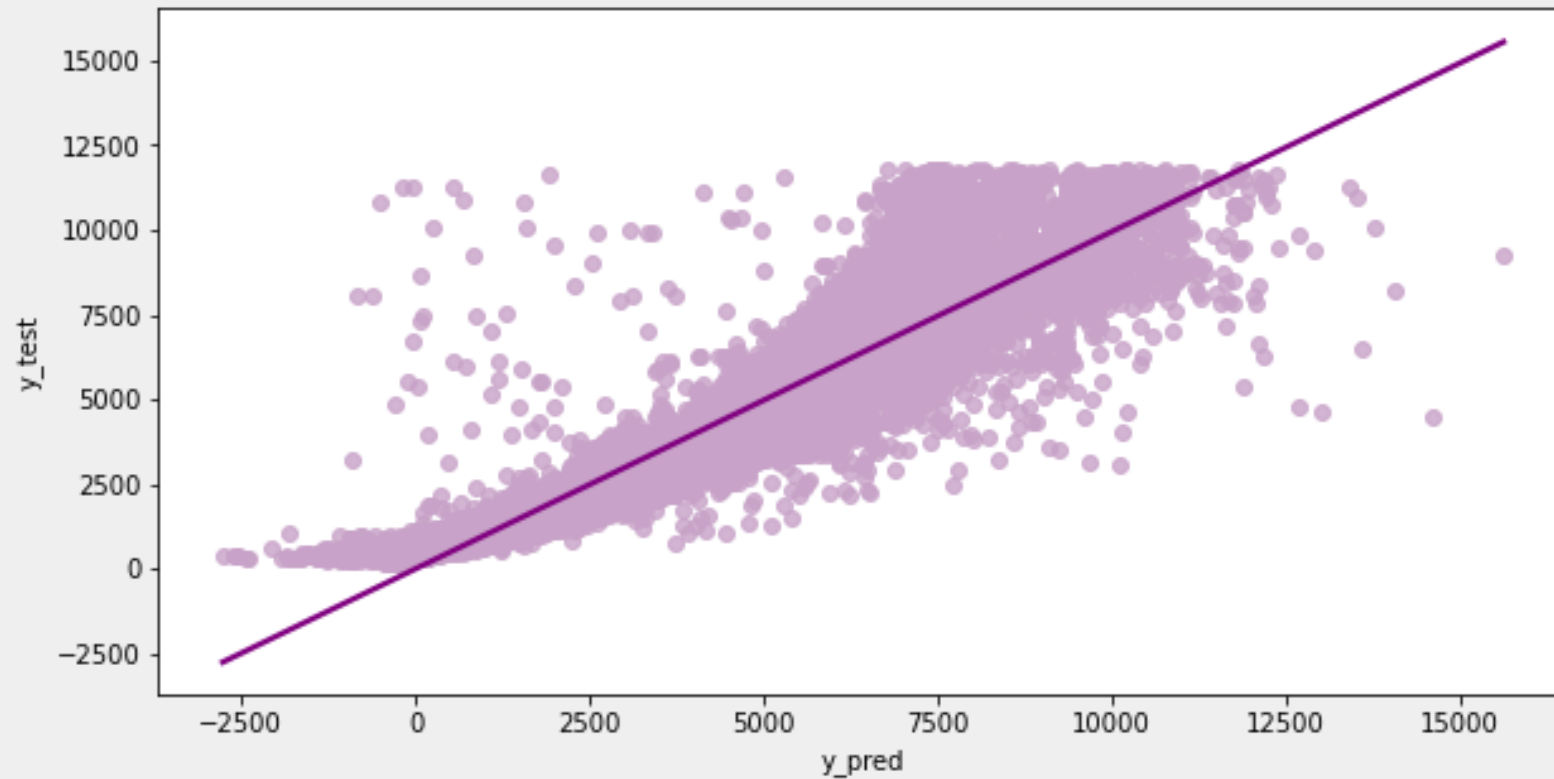
## Dropped rows $\approx$ 2%

- duplicates





# Feature Engineering



## Residual Plot

Removing outliers improved our model by 20%



## Accuracy

MAE: 440 AUD  
R2/Adjusted R2: 0.90

Our model is able to account approximately 90% of the variation in sales price of a diamond and it is able to predict the sales price with the Mean Absolute Error of 440 AUD.

# Feature Engineering

---



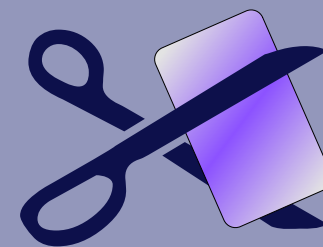
## Carat

a carat equals the weight of 0.2 grams



## Clarity

purity, rarity and % blemishes, inclusions



## Cut

facets, symmetry, dimensions and reflective qualities



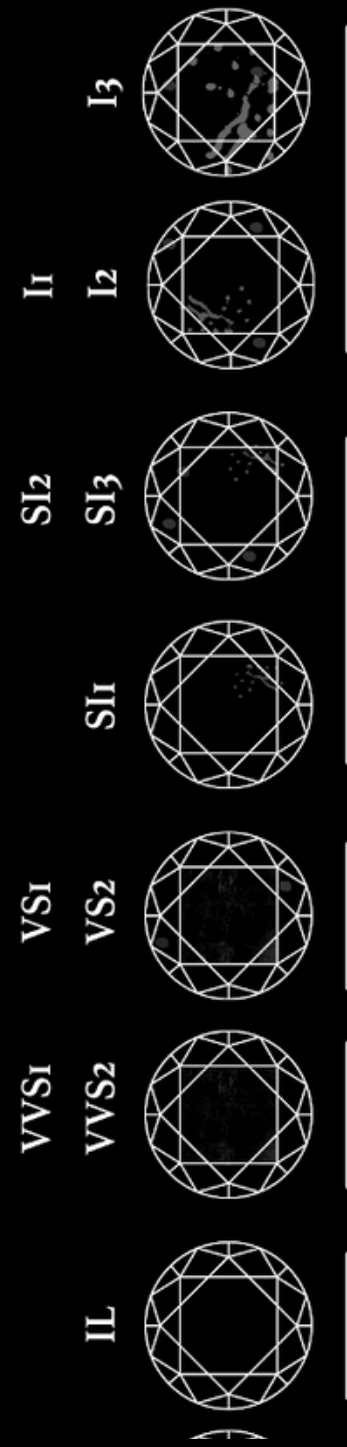
## Color

poor color grade (M) will have a visible yellow tint





# Feature Engineering: clarity



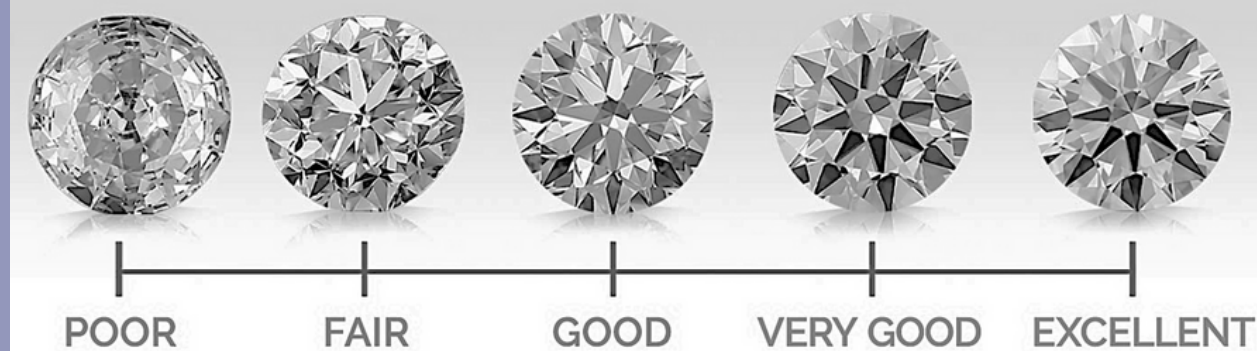
Clarity	Carat											Avg. Total Sales Price	
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0		
I1	4,624	686	1,309	2,403	3,472	5,268	6,737	8,773	8,182	8,594	9,971	378	11,503
I2	5,865	886	1,165	1,887	2,553	3,826	5,366	7,709	8,257				
I3		378	812	1,399	2,126	2,778	3,080	4,456					
SI1	1,059	847	1,859	3,282	4,636	7,066	8,587	9,034	9,141	10,518	9,443		
SI2	3,067	755	1,593	3,026	4,086	6,177	7,877	9,105	9,259	9,250	9,476		
VS1	543	928	2,279	3,763	5,655	8,137	8,430	9,335	9,263	8,734			
VS2	695	901	2,184	3,729	5,237	7,972	8,225	9,338	9,334	10,166	9,573		
VVS1	639	1,020	2,498	4,199	6,932	7,952	8,384	9,471	9,757		11,503		
VVS2	606	953	2,355	3,905	6,221	7,890	8,387	9,414	9,556	10,375	10,108		
IF	666	1,187	2,886	4,742	7,175	8,390	8,706	9,336	10,275				

Clarity	Color										Avg. Total Sales Price	
	D	E	F	G	H	I	J	K	L	M		
I1	2,892	2,727	2,973	3,104	3,278	3,592	2,898	3,222	2,994	2,901	1,161	3,592
I2	2,477	2,713	2,456	2,721	3,295	2,986	2,483	2,926	1,595	1,794		
I3	1,161	1,576	2,223	1,878	2,147	1,388	1,673		1,548	1,510		
SI1	2,393	2,073	2,250	2,490	2,591	2,810	2,582	2,329	2,423	2,344		
SI2	2,583	2,241	2,500	2,947	3,270	2,980	3,077	2,609	2,590	2,729		
VS1	2,049	1,959	2,117	2,654	2,878	2,529	2,271	2,213	2,166	2,286		
VS2	2,409	2,160	2,393	2,764	2,843	2,417	2,478	2,143	2,437	2,451		
VVS1	1,973	1,783	1,836	2,263	2,447	2,204	2,073	2,009	2,203	2,362		
VVS2	1,806	1,736	1,834	2,389	2,594	2,448	2,308	1,969	2,300	2,565		
IF	2,693	1,800	2,006	2,372	2,816	2,377	2,401	2,458	2,989	3,210		

# Feature Engineering: cut

Clarity	Cut				Total Sales Price	
	Fair	Good	Very Good	Excellent		
I1		5.011	4.235.819	8.853.504	506	43.742.901
I2		1.066	420.735	1.057.862		
I3			34.456	43.281		
IF			1.107.186	13.494.420		
SI1		12.297	12.540.172	41.340.031		
SI2	1.491	24.296	12.216.538	35.314.350		
VS1		3.365	11.650.706	38.065.289		
VS2		506	13.020.066	43.742.901		
VVS1			6.078.084	31.075.921		
VVS2			8.706.247	29.897.789		

Color	Cut				Avg. Total Sales Price	
	Fair	Good	Very Good	Excellent		
Null	1,491	2,120	3,967	3,651	1,491	10,307
D			2,366	2,062		
E		6,735	2,275	1,865		
F		10,307	2,480	2,019		
G			2,633	2,513		
H		6,606	2,714	2,720		
I			2,511	2,478		
J		4,844	2,279	2,400		
K		3,212	1,955	2,221		
L			2,009	2,469		
M			2,096	2,570		

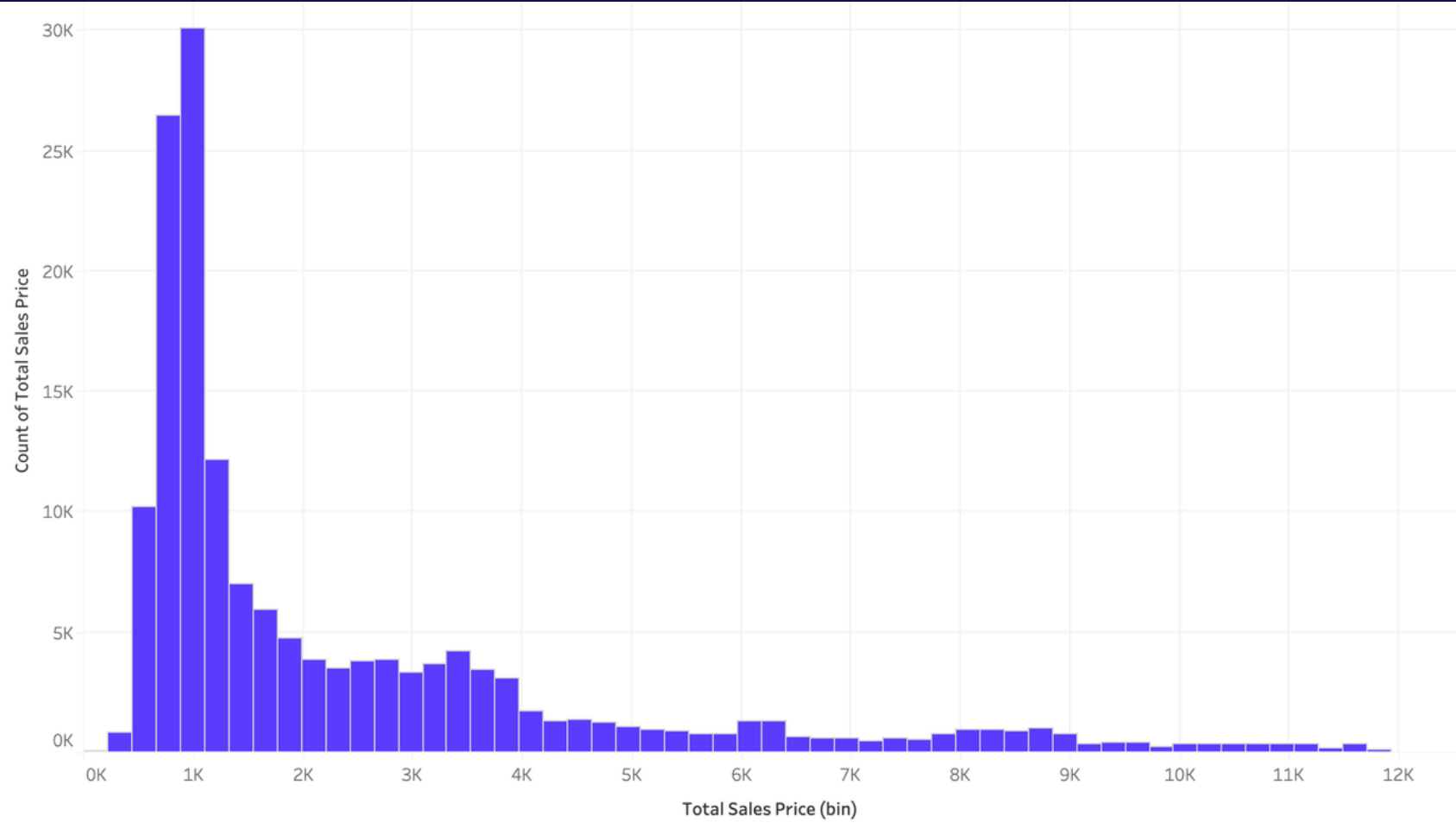
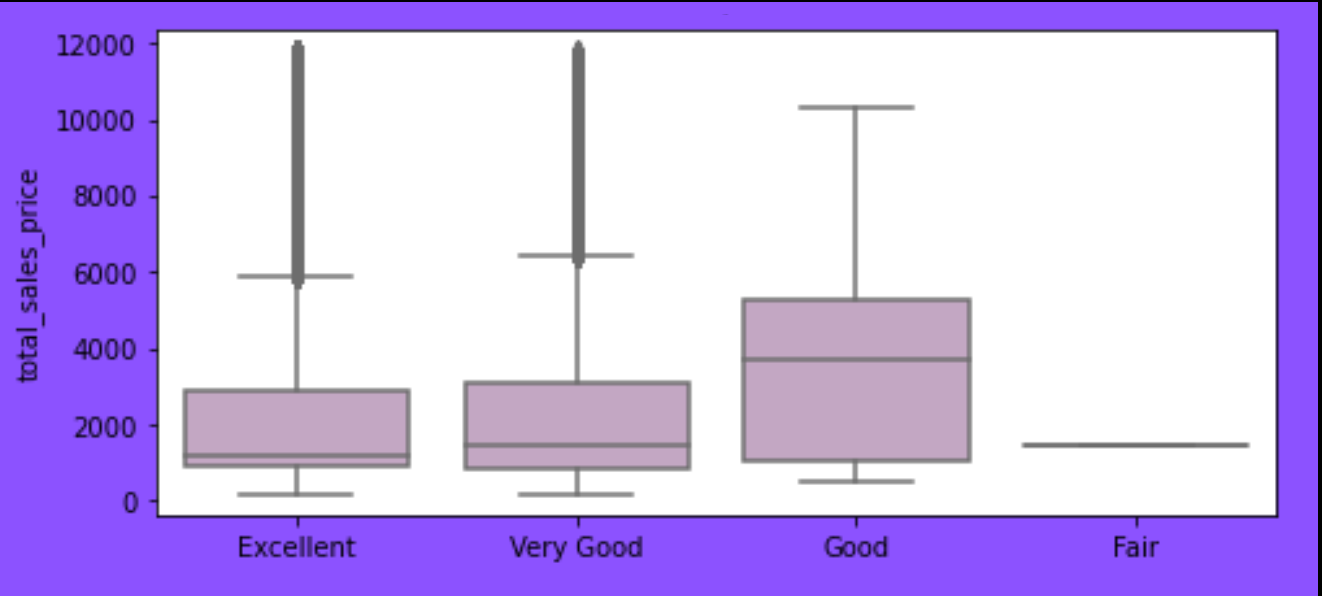




# How Much Does A Diamond Cost?

Color	Carat											Avg. Total Sales Price
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	
L		579	1,086	2,082	3,225	4,411	5,985	8,154	8,794	9,697	9,498	
M		510	914	1,884	2,890	4,006	5,424	6,964	7,830	10,488	8,660	

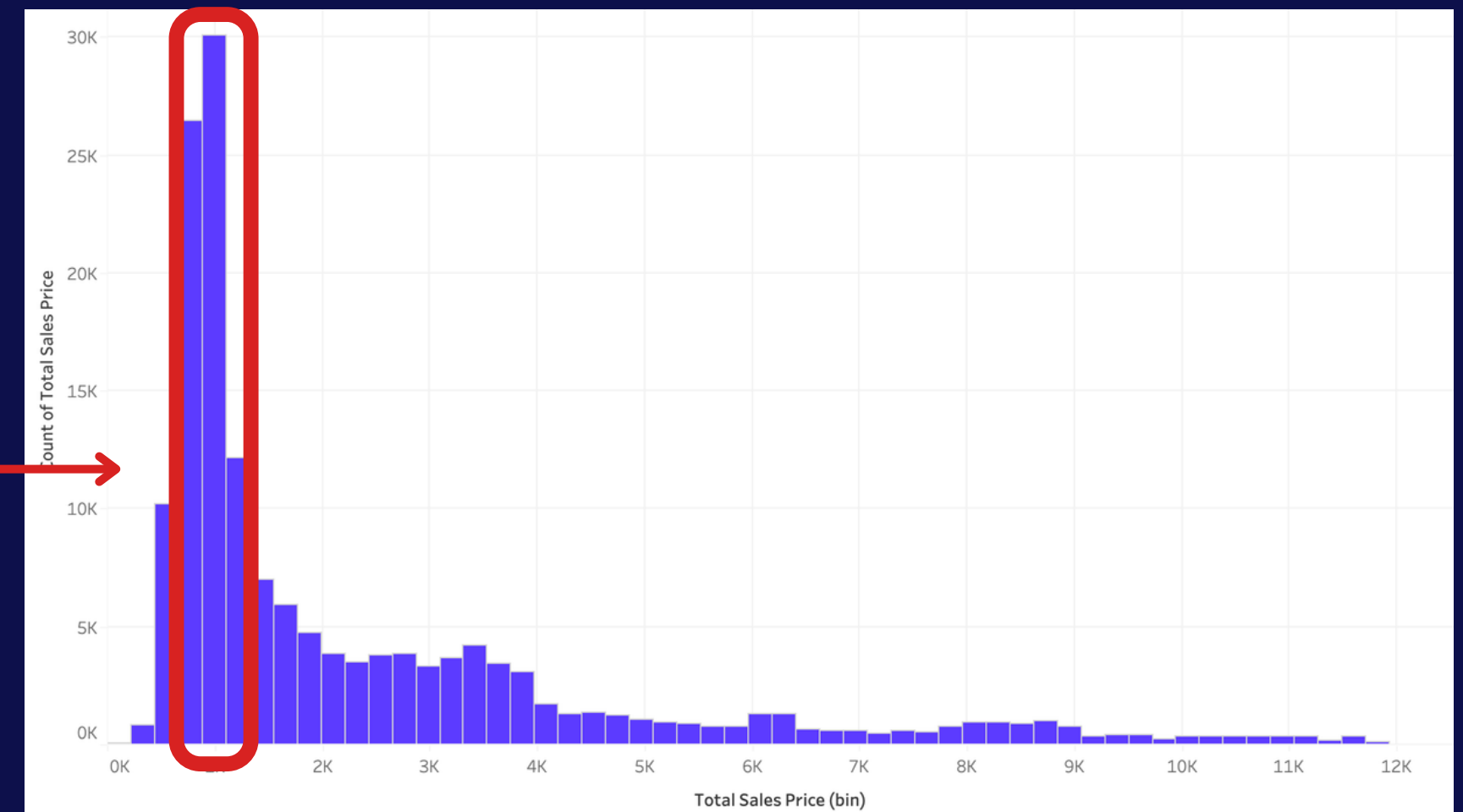
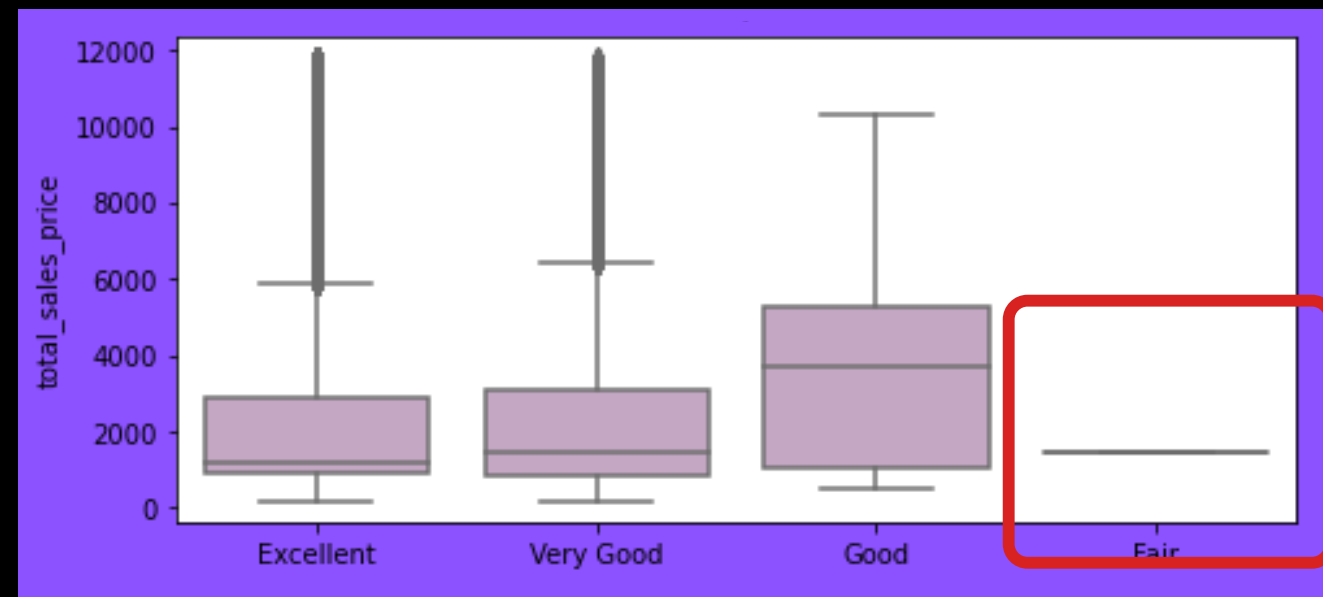
Clarity	Carat											Avg. Total Sales Price
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	
I1	4,624	686	1,309	2,403	3,472	5,268	6,737	8,773	8,182	8,594	9,971	378
I2	5,865	886	1,165	1,887	2,553	3,826	5,366	7,709	8,257			11,503



# How Much Does A Diamond Cost?

Color	Carat											Avg. Total Sales Price
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	
L		579	1,086	2,082	3,225	4,411	5,985	8,154	8,794	9,697	9,498	
M		510	914	1,884	2,890	4,006	5,424	6,964	7,830	10,488	8,660	

Clarity	Carat											Avg. Total Sales Price
	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	
I1	4,624	686	1,309	2,403	3,472	5,268	6,737	8,773	8,182	8,594	9,971	378
I2	5,865	886	1,165	1,887	2,553	3,826	5,366	7,709	8,257			11,503





# Conclusions

01

Diamonds price correlates with

Carat



Clarity



Cut



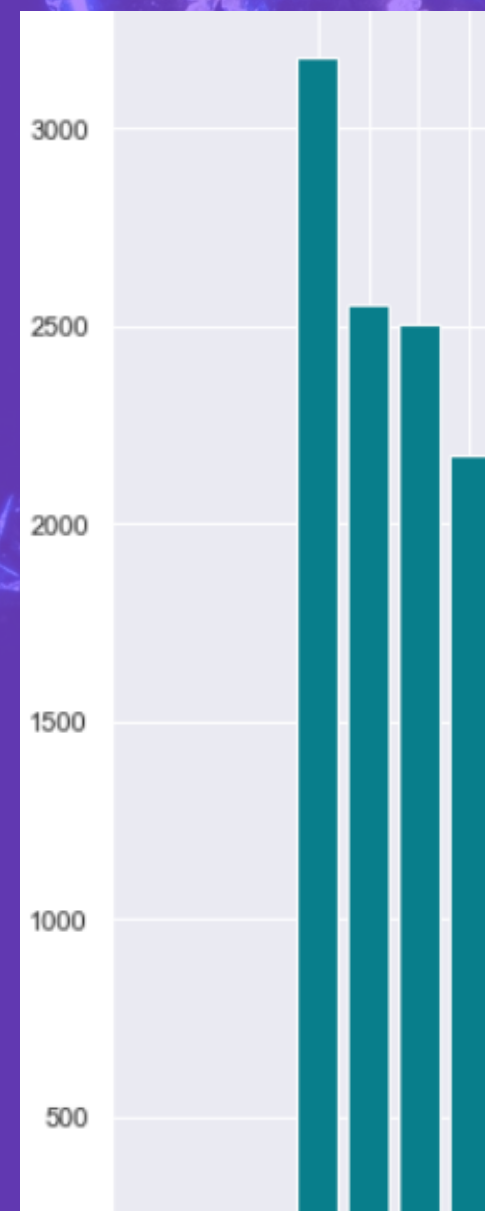
Color



02

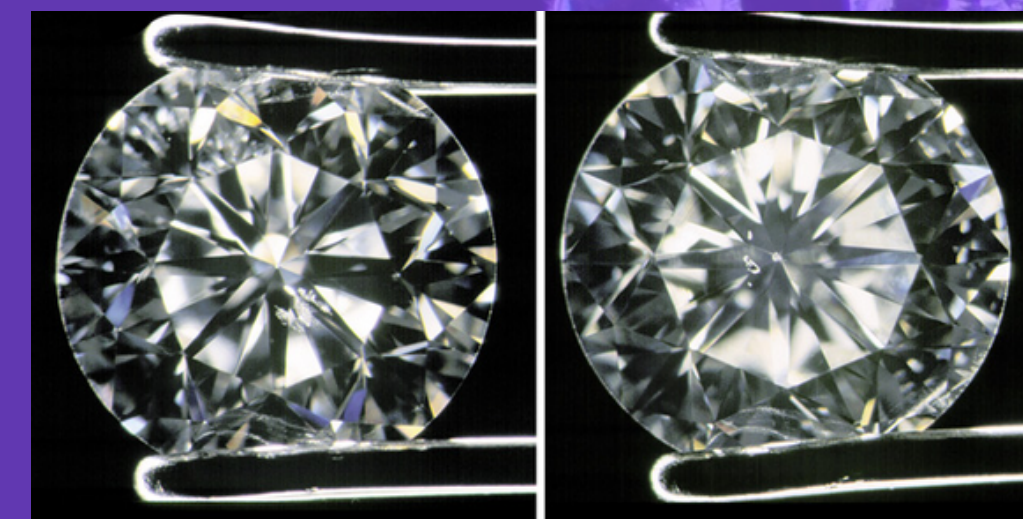
People demand affect the prices more than diamonds quality

Features Importance



03

Prices sharply increase for bigger carat weight than for a better cut, color quality or clarity.



Excellent cut

Fair cut



# Conclusions

01

Diamonds price correlates with

Carat



Clarity



Cut



Color



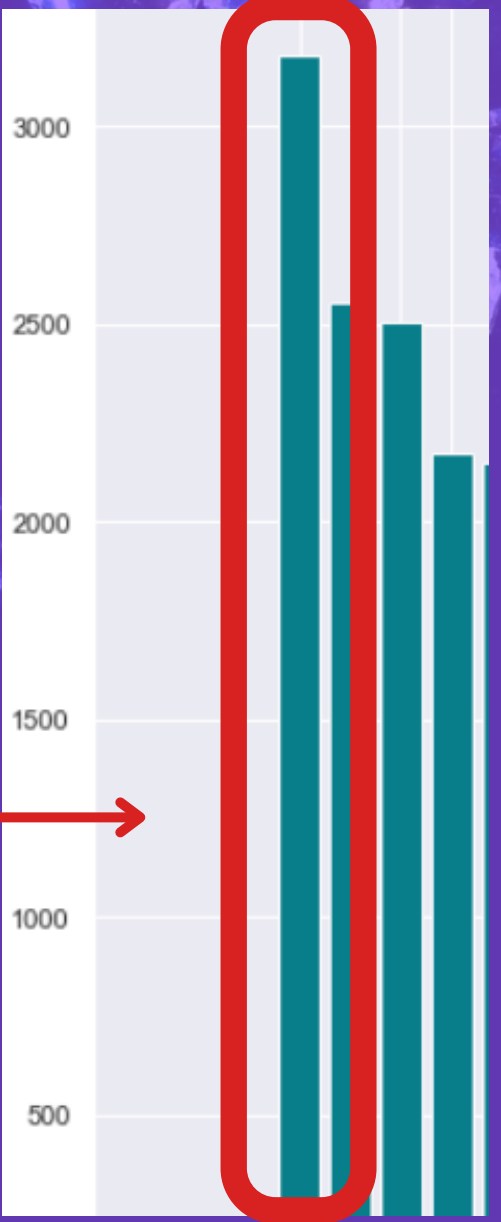
02

People demand affect the prices more than diamonds quality

Fair cut

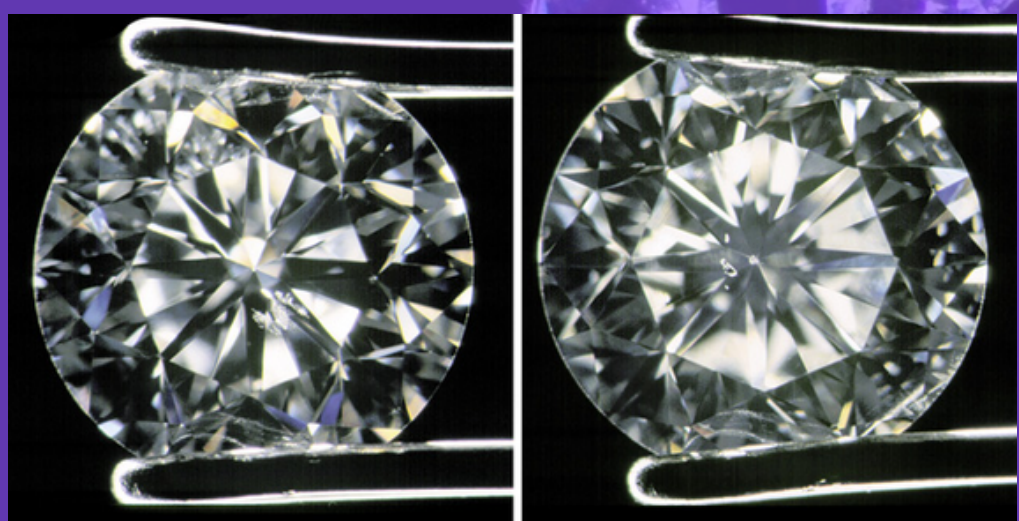


Features Importance



03

Prices sharply increase for bigger carat weight than for a better cut, color quality or clarity.



Excellent cut

Fair cut



# References

---

<https://australiandiamondimporters.com.au/>

<https://www.kaggle.com/datasets/enashed/diamond-prices>

<https://www.gemsociety.org/article/diamond-rarity-quality-and-cost/>

[https://www.jamesallen.com/education/diamonds/4cs-carat/?a\\_aid=1212](https://www.jamesallen.com/education/diamonds/4cs-carat/?a_aid=1212)

<https://www.gia.edu/gia-about/4cs-cut>

