

DIAMONDS

Becerra H., Fumagalli V., Gurtan A., Pandolfi R.

Statistics - BABD 2023



THE DATASET

The diamonds dataset.

CUT

CLARITY

DEPTH (MM)

COLOR

LENGTH (MM)

DEPTH (%)

CARAT

WIDTH (MM)

TABLE

PRICE (US
DOLLARS)

WHY AND WHO?

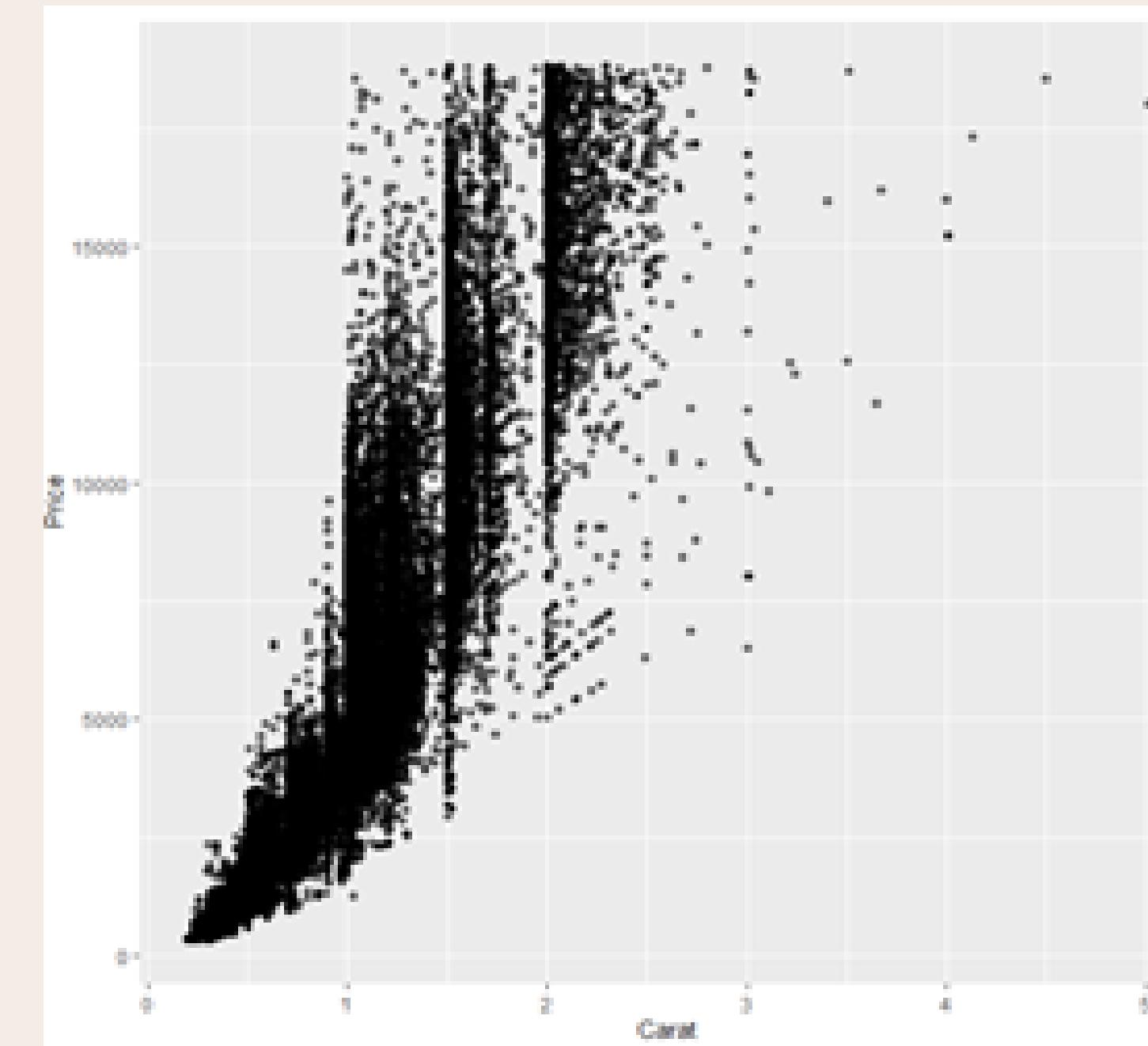
- Jewelry diamond market valued at \$249 billion in 2021
- New trends, new models
- Jewelry companies, wholesalers, buyers
- Associations: Confindustria Federorafi; CIBJO – The World Jewellery Confederation

OUR GOALS

- What are the most common features of diamonds?
- Which features should I expect from the highest quality diamonds?
- How much should my diamond be priced?

OVERVIEW

- Descriptive analysis
- Max. price 18823 -> 3 carats



DESCRIPTIVE ANALYSIS

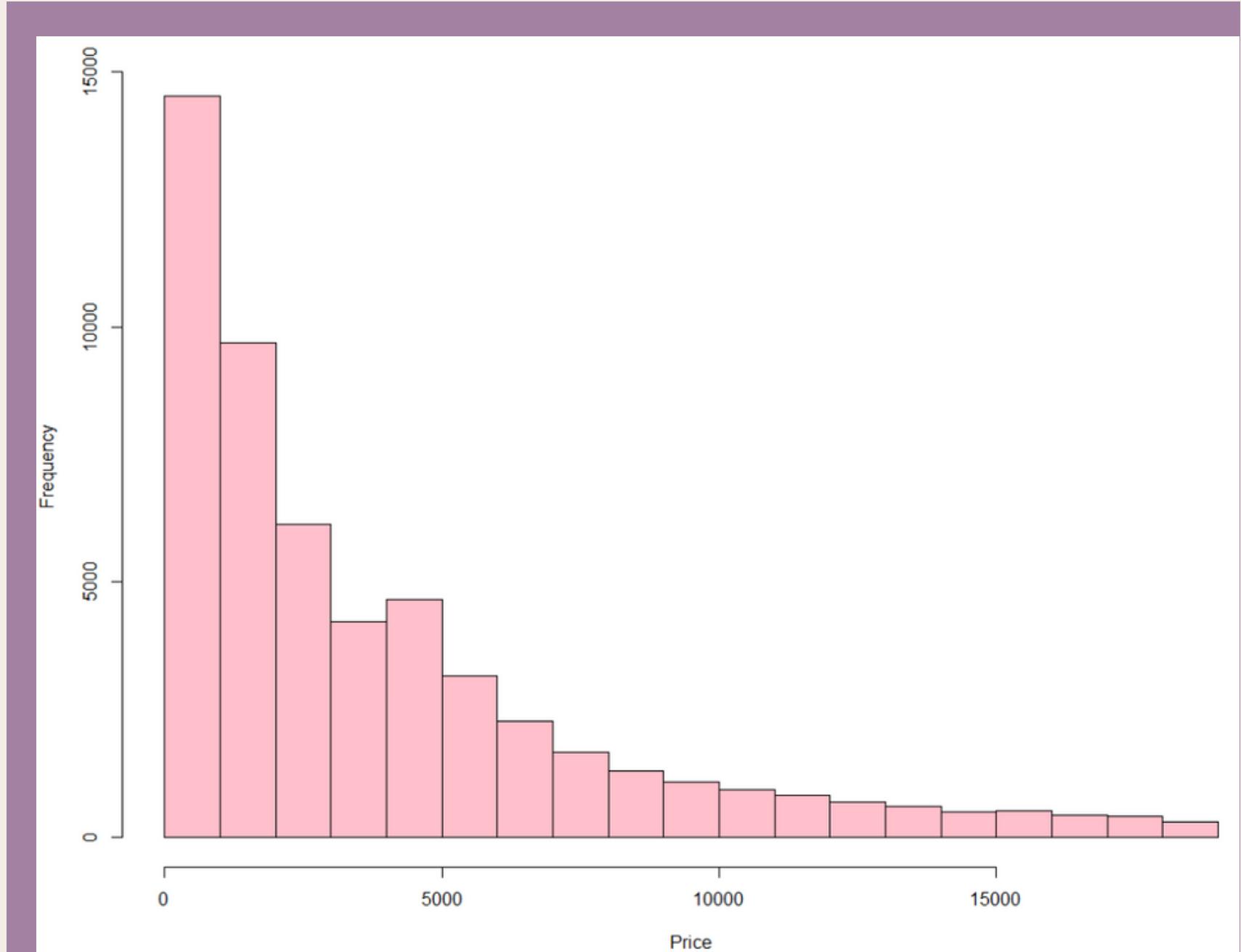
carat	cut	color	clarity	depth	table
Min. :0.2000	Fair : 1600	D: 6774	SI1 :13065	Min. :43.00	Min. :43.00
1st Qu.:0.4000	Good : 4903	E: 9796	VS2 :12257	1st Qu.:61.00	1st Qu.:56.00
Median :0.7000	Very Good:12080	F: 9541	SI2 : 9182	Median :61.80	Median :57.00
Mean :0.7964	Premium :13778	G:11291	VS1 : 8171	Mean :61.75	Mean :57.46
3rd Qu.:1.0400	Ideal :21547	H: 8298	VVS2 : 5066	3rd Qu.:62.50	3rd Qu.:59.00
Max. :3.0000		I: 5409	VVS1 : 3655	Max. :79.00	Max. :95.00
		J: 2799	(Other): 2512		
price	x	y	z		
Min. : 326	Min. :0.000	Min. : 0.000	Min. : 0.000		
1st Qu.: 949	1st Qu.:4.710	1st Qu.: 4.720	1st Qu.: 2.910		
Median : 2400	Median :5.700	Median : 5.710	Median : 3.520		
Mean : 3926	Mean :5.729	Mean : 5.732	Mean : 3.537		
3rd Qu.: 5316	3rd Qu.:6.540	3rd Qu.: 6.540	3rd Qu.: 4.030		
Max. :18823	Max. :9.420	Max. :58.900	Max. :31.800		

*x: length (mm)

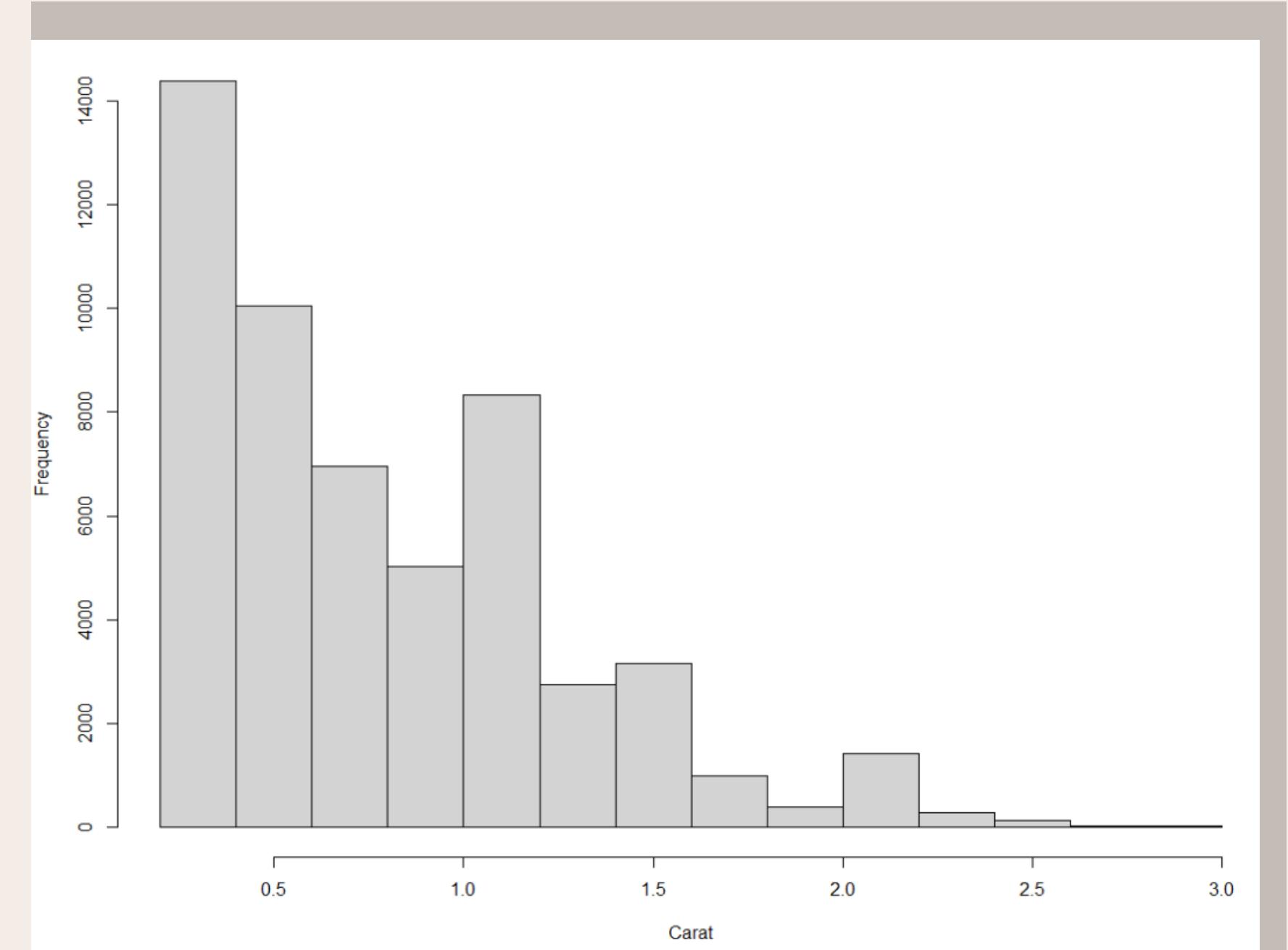
y: width (mm)

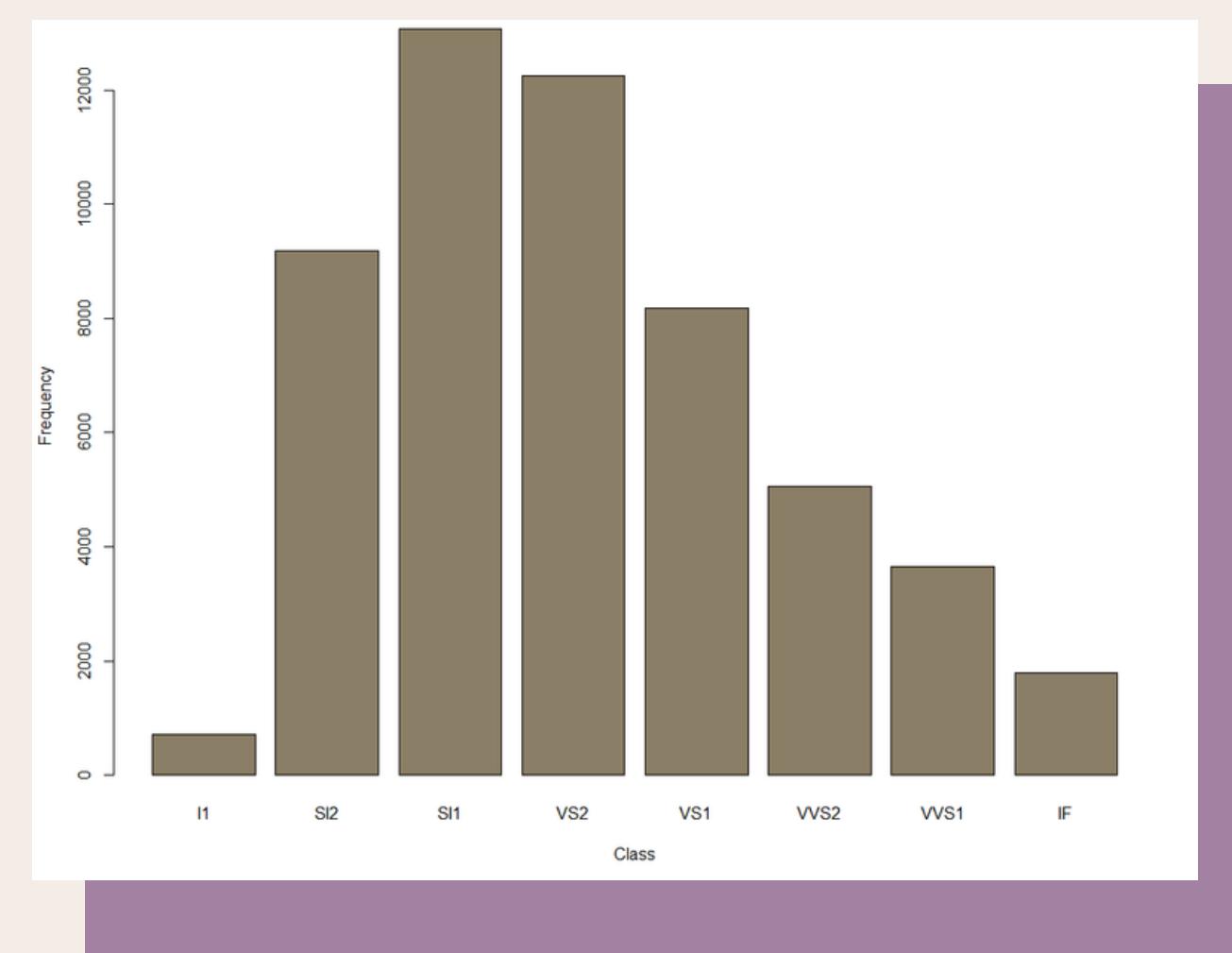
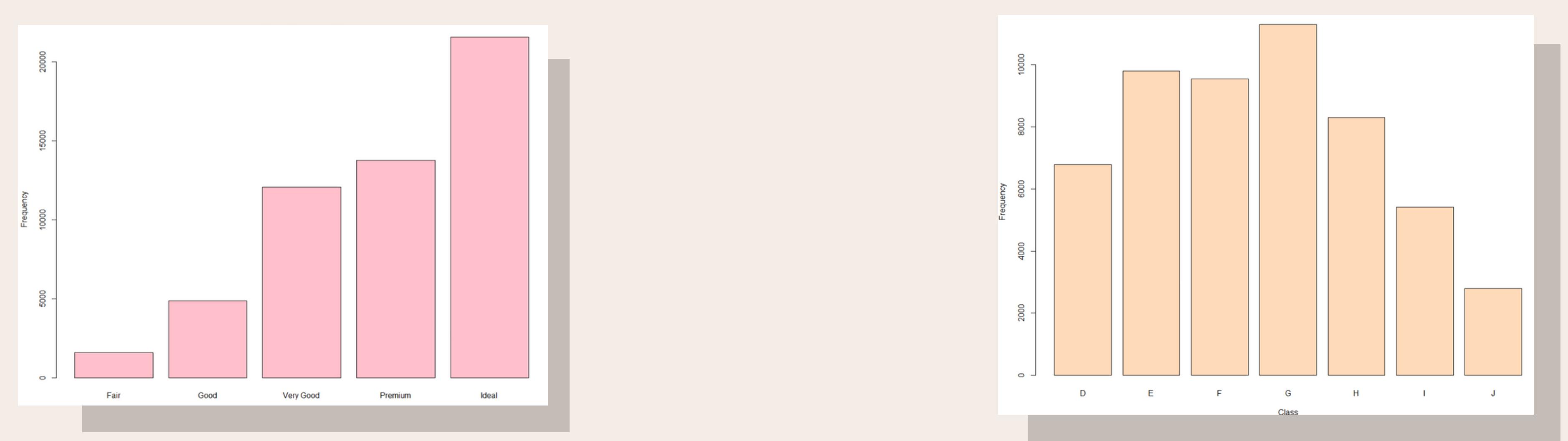
z: depth (mm)

Price



Carat





IDENKIT: THE MOST PRECIOUS DIAMOND

INTERNALLY
FLAWLESS

Price 2865

Carat 0.505

Color G

Cut Ideal



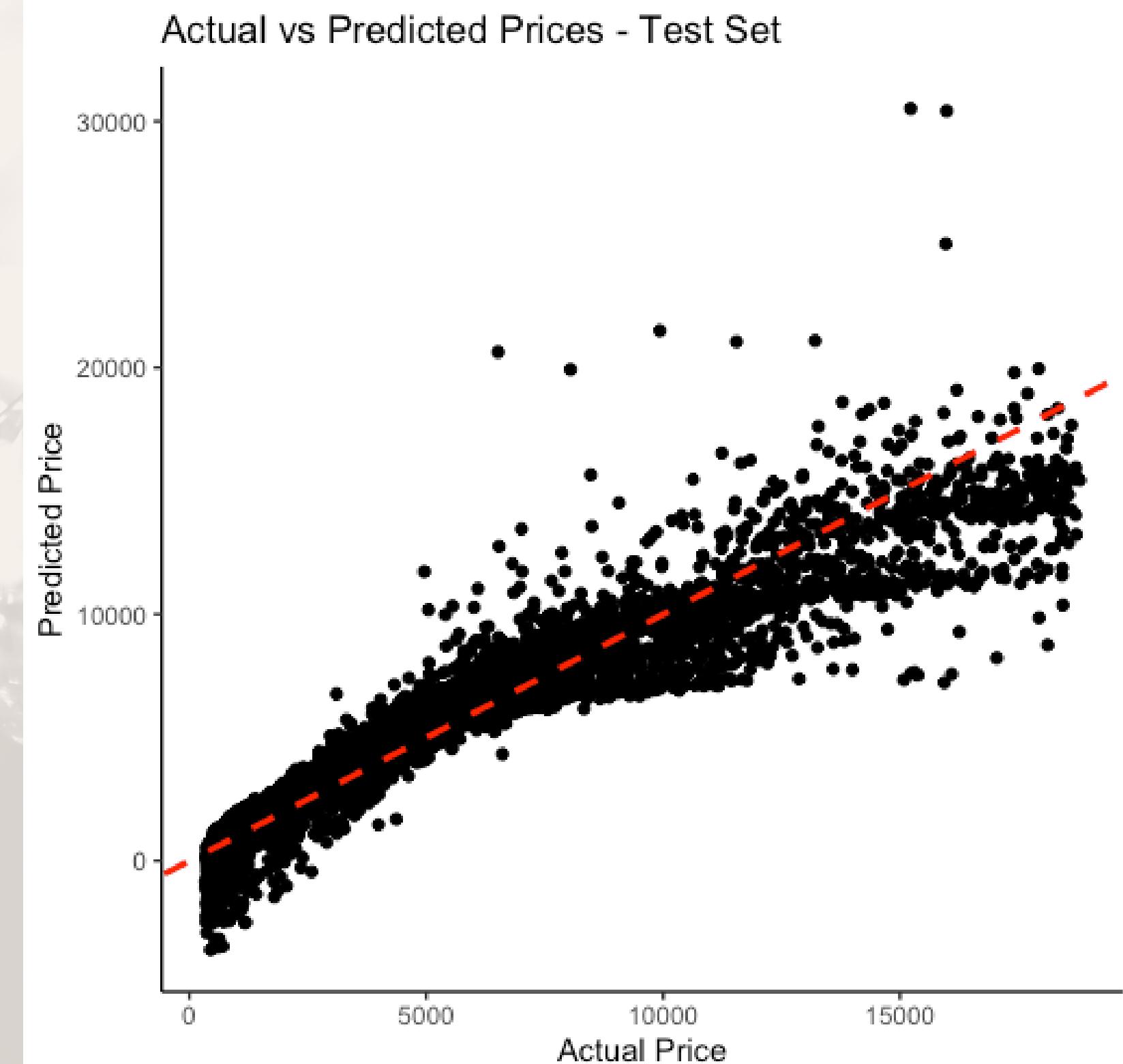
PRICE PREDICTION

- Create **Dummy variables** for the **categorical** variables 'cut', 'color', and 'clarity'
- Selecting the **predictors** and **response** variables and splitting the data into **training** and **testing** sets
- Fit a **Lasso Regression** model to the training data, using cross-validation to tune the Lambda value
- Refitting the model using the best lambda value and extracting the coefficients
- Calculating the root mean squared error (**RMSE**) on both the **training** and **testing** data, as well as using **k-fold** cross-validation to **estimate** the test error, then compare them

Training Set --> RMSE : 1127.68 , ADJ R2: 92,04%

Test Set --> RMSE: 1139.4, ADJ R2: 91,68%

RMSE estimated with k fold cross-validation: 1130.99





WHY US?

- Diamonds are priced by the 4c's, which misses a standard pricing for other traits of a diamond other than cut, color, clarity and carat.
- Our priority is transparency: if you buy our model, both the buyer and the seller will base their price on a clear standardized price, with lesser risk of loss.
- High prediction accuracy model

LIMITATIONS

- Dataset has an upper limit on diamond prices
- Dataset only has colorless and near-colorless ranges
- PCA
- Linear model
- Variable correlations



THANK YOU!

REFERENCES

International Gem Society. (2022). The Fifth C: What Determines Diamond Cost? - International Gem Society. [online] Available at: <https://www.gemsociety.org/article/what-determines-diamond-cost/>

SkyQuest Technology Consulting Pvt. Ltd (2022). Global Jewelry Market to Reach \$443 Billion by 2028 | Diamond Jewelry Demand Surpassed Sales Worth \$94 billion in 2021 | SkyQuest Technology. [online] GlobeNewswire News Room. Available at: <https://www.globenewswire.com/en/news-release/2022/11/16/2557366/0/en/Global-Jewelry-Market-to-Reach-443-Billion-by-2028-Diamond-Jewelry-Demand-Surpassed-Sales-Worth-94-billion-in-2021-SkyQuest-Technology.html>

Skyquestt.com. (2017). SkyQuest Technology. [online] Available at: <https://skyquestt.com/>

Swati Khedekar (2022). Data Analysis on Diamonds Dataset. [online] Kaggle.com. Available at: <https://www.kaggle.com/datasets/swatikhedekar/price-prediction-of-diamond>

