Brides

Regression Model for

# DIAMONDS PRICES DATASET

This material is made only for learning material
**PPT & Algoritm by Handoko Pramulyo**
**Batch : FTDS-014**

06/Sept/2022

Since there are quite a number of parameters that can determine the price of diamonds, it will be difficult for retail business owners when they want to release the new prices for any specs of diamonds to the public. In some cases, business owners also find it hard to buy diamonds at the right price from the principal & manufacture of the diamonds. 🔑

People give advice about taking the best diamond price on the market by trust your own eyes on what you like, ask for quality, and negotiate for the best price. **Obviously, this is why estimating diamonds prices turn up to be so hard in real life.**

Quora    The New York Times    🩶

**Problem Statement & Objective**

So, by **making a Machine Learning Regression Model**, this subjectivity price estimation with a large dollar impact value can be an alternative to determining the diamond price properly without the need to hire a diamond price estimator.
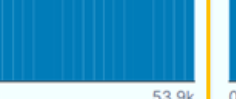
# Diamonds Prices

Learning Diamonds Data Exploration

https://www.kaggle.com/datasets/nancyalaswad90/diamonds-prices

**MODEL YANG DIGUNAKAN**
**XGBoost for Machine Learning**
is one of the most well known
algorithms in statistics & ML
machinelearningmastery.com

*dmlc*
**XGBoost**

| # Index | # carat float64 | △ cut object | △ color object | △ clarity object | # depth float64 | # table float64 | # price int64 | # x float64 | # y float64 |
|---|---|---|---|---|---|---|---|---|---|
| | | Ideal 40% | G 21% | SI1 24% | | | | | |
| | | Premium 26% | E 18% | VS2 23% | | | | | |
| | | Other (18599) 34% | Other (32852) 61% | Other (28617) 53% | | | | | |
| 1 ... 53.9k | 0.2 ... 5.01 | | | | 43 ... 79 | 43 ... 95 | 326 ... 18.8k | 0 ... 10.7 | 0 ... 58.9 |
| 1 | 0.23 | Ideal | E | SI2 | 61.5 | 55 | 326 | 3.95 | 3.98 |
| 2 | 0.21 | Premium | E | SI1 | 59.8 | 61 | 326 | 3.89 | 3.84 |
| 3 | 0.23 | Good | E | VS1 | 56.9 | 65 | 327 | 4.05 | 4.07 |
| 4 | 0.29 | Premium | I | VS2 | 62.4 | 58 | 334 | 4.2 | 4.23 |
| 5 | 0.31 | Good | J | SI2 | 63.3 | 58 | 335 | 4.34 | 4.35 |
| 6 | 0.24 | Very Good | J | VVS2 | 62.8 | 57 | 336 | 3.94 | 3.96 |
| 7 | 0.24 | Very Good | I | VVS1 | 62.3 | 57 | 336 | 3.95 | 3.98 |
| 8 | 0.26 | Very Good | H | SI1 | 61.9 | 55 | 337 | 4.07 | 4.11 |
| 9 | 0.22 | Fair | E | VS2 | 65.1 | 61 | 337 | 3.87 | 3.78 |
| 10 | 0.23 | Very Good | H | VS1 | 59.4 | 61 | 338 | 4 | 4.05 |
| 11 | 0.3 | Good | J | SI1 | 64 | 55 | 339 | 4.25 | 4.28 |
| 12 | 0.23 | Ideal | J | VS1 | 62.8 | 56 | 340 | 3.93 | 3.9 |
| 13 | 0.22 | Premium | F | SI1 | 60.4 | 61 | 342 | 3.88 | 3.84 |

CUSTOMMADE    GIA   Gemological Institute of America Inc
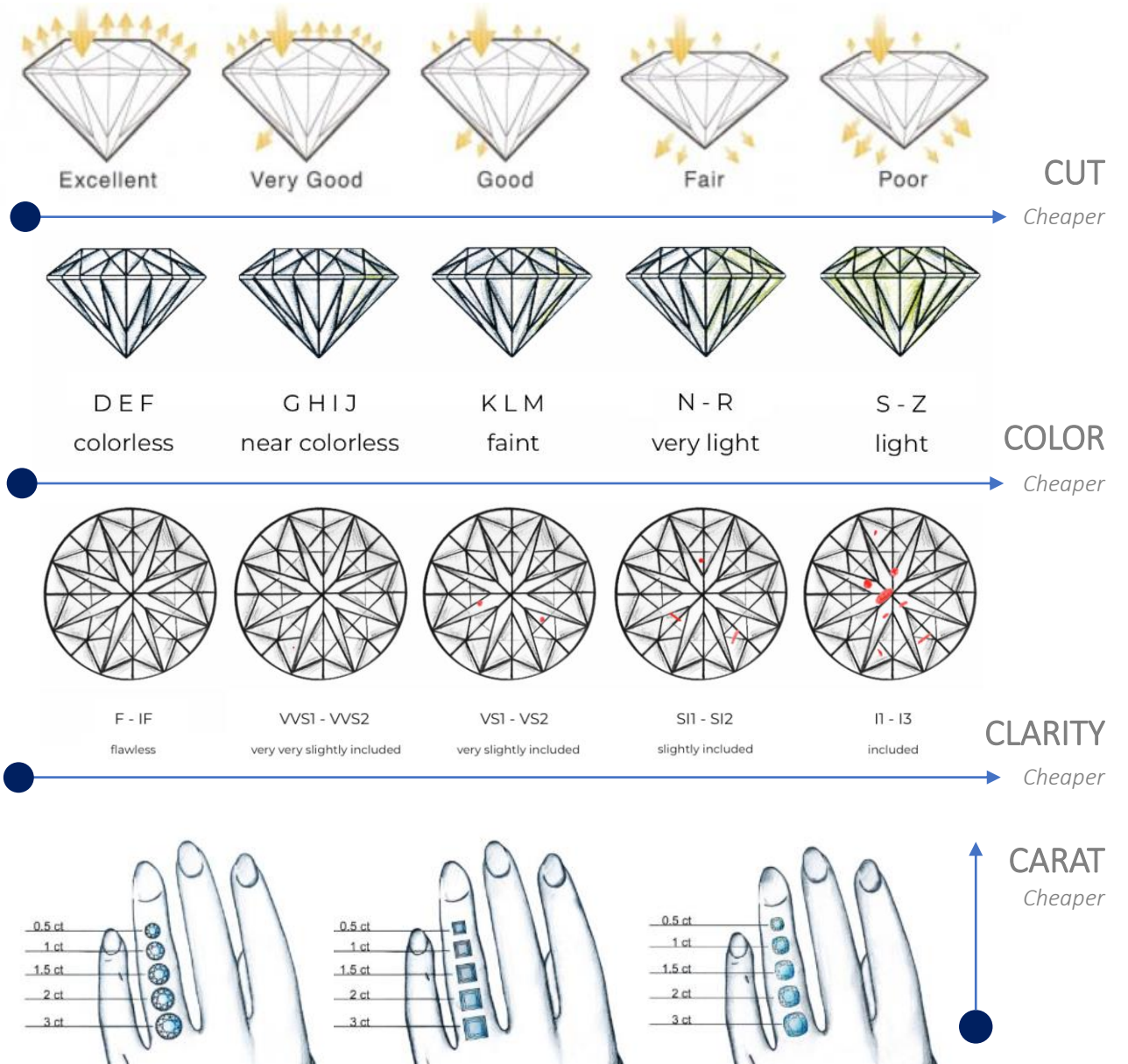
**DIAMOND KNOWLEDGE**
Learn how diamonds are graded on cut, color, clarity and carat weight to get a perfect center stone at a great value.
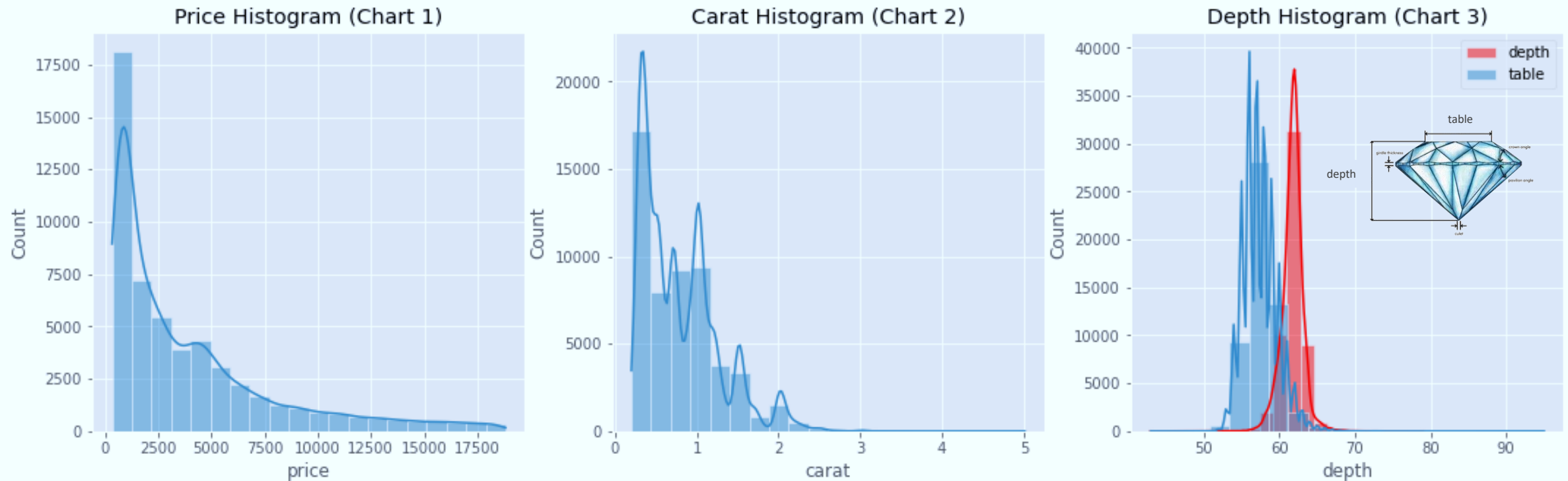


The Diamond Store -

**WHAT IS 4C DIAMOND?**
4C refers to the global standard for the four main characteristics of diamonds: **Cut, Colour, Clarity & Carat.** These properties interact to produce the beauty and luster that makes diamonds so attractive.

| Excellent | Very Good | Good | Fair | Poor | CUT |
|---|---|---|---|---|---|
| | | | | | Cheaper |

| D E F colorless | G H I J near colorless | K L M faint | N - R very light | S - Z light | COLOR |
|---|---|---|---|---|---|
| | | | | | Cheaper |

| F - IF flawless | VVS1 - VVS2 very very slightly included | VS1 - VS2 very slightly included | SI1 - SI2 slightly included | I1 - I3 included | CLARITY |
|---|---|---|---|---|---|
| | | | | | Cheaper |

CARAT
Cheaper

0.5 ct
1 ct
1.5 ct
2 ct
3 ct

- **In Chart 1, it can be seen that the most sold number of diamonds (the most sales) are diamonds with prices in the range of $ ~ 0 to 2000** which makes sense because in terms of price it is much more affordable for buyers. However, it is very unexpected that there are several diamonds with very expensive prices that still have a market in the market, such as collectors, billionaires, auctions, etc.
- **In Chart 2, it can be seen that the number of carat ranges from ~ 0 to 1 and is the most common diamond** found in the market. 
- As for chart 3, most diamond manufacturers generally make **diamonds with a proportional dimension ratio**, both in terms of table and depth. However, some table sizes and depths are still far from ideal (most probably due to limitations in machine technology during production).
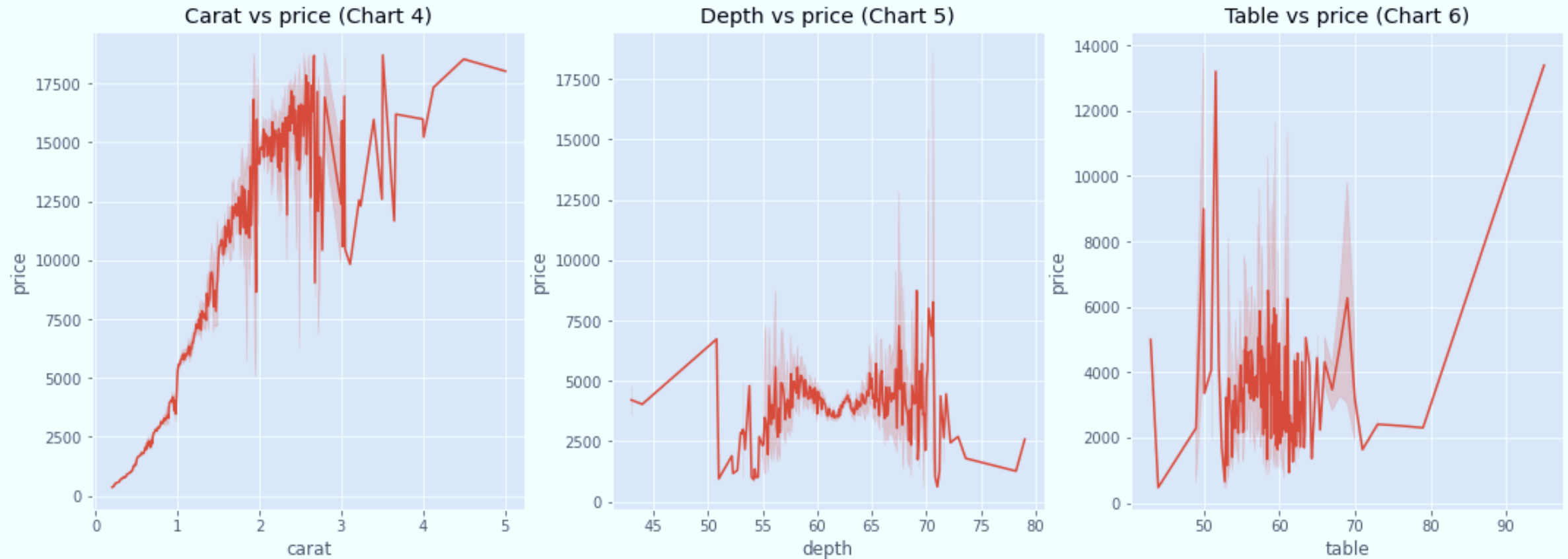
- **It is common knowledge that carat is always the main parameter in determining the price of diamonds** (see chart 4), the higher the carat of a diamond, the higher the price of course (the same concept with gold).However, something unique is seen in the carat range 3 to 5, the price mostly remains the prices that are not much different from diamonds with carat ~ 2 to 3.
- Chart 5 and 6, **depth and table are not significantly correlated with price**, chart show that there is less + or - trend between the three features.

Cut vs price (Chart 7)

Color vs price (Chart 8)

Clarity vs price (Chart 9)
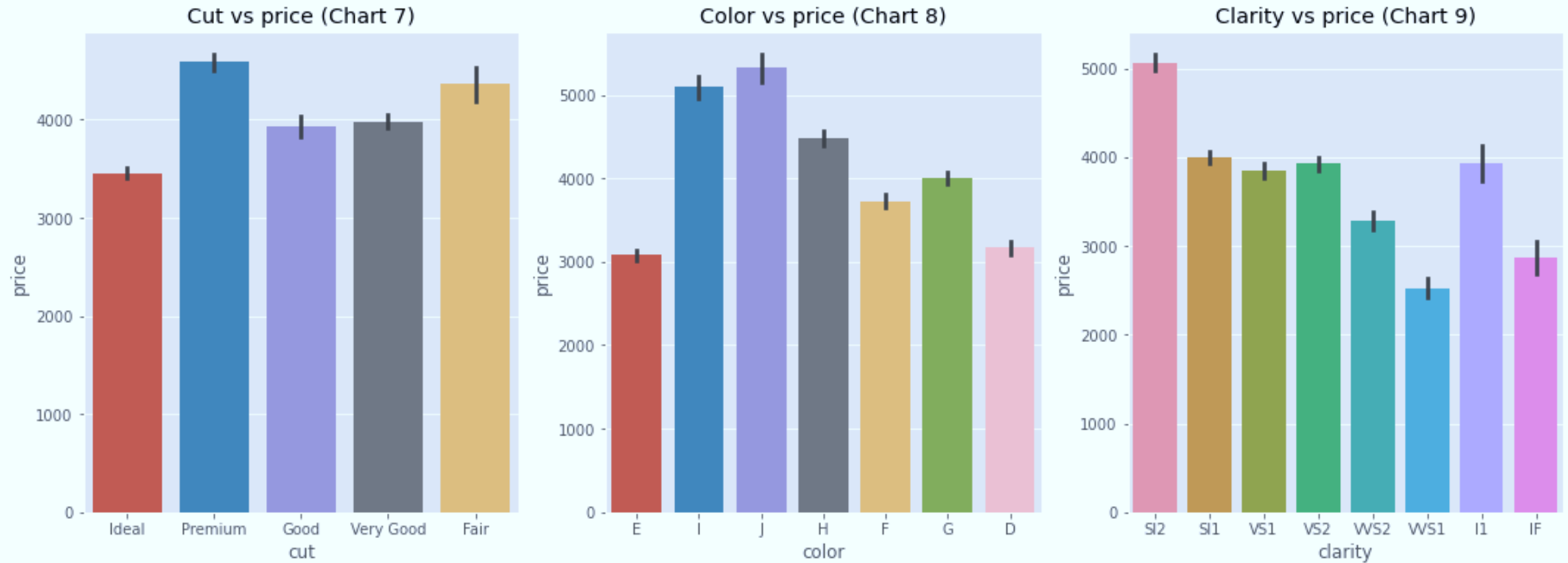
- **Premium cut** diamond turned out to be the diamond with the **highest price.**
- As for color (see chart 8), diamonds with colors in the **range I to H become** diamonds with the highest prices.
- In chart 9, **clarity at the SI2 and SI1 grade levels is the most expensive**.

GIA®

The quality of the 4C diamond does not always depend on the quality of the diamonds production process itself, but also depends on the source of the quality of the mining diamond (somehow miners found diamond with high carat but poor clarity and dimension, or miners found diamond with excellent color, clarity, and awesome size, but unfortunately the carat is small and the cutting process fails, etc)
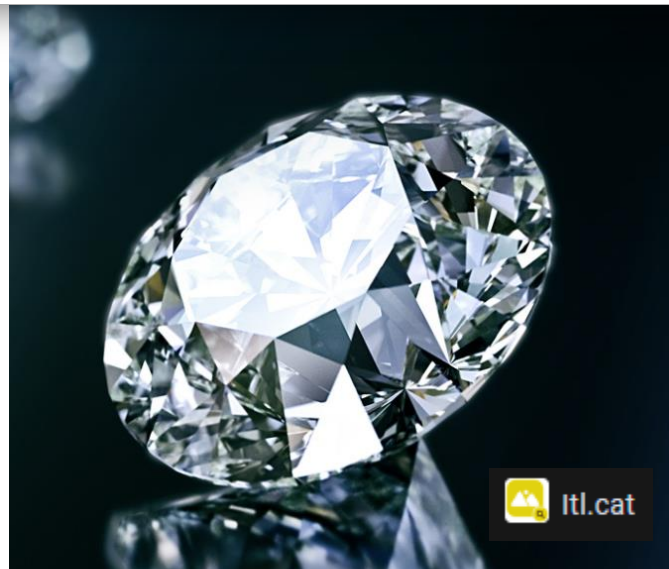
## dmlc XGBoost

Regression Model For
**Diamonds Prices Quality ~**

**~98%**

ERROR AROUND
~ $268 & $305


ltl.cat

## Model Deployment



**Aplikasi Estimasi Diamond Price**

carat
0 — 1 — 5

cut
Fair

color
J

clarity
I1

depth
43 — 55 — 80

table
43 — 60 — 95

x
0 — 5 — 11

y
0 — 10 — 60

z
0 — 15 — 32

**9295.93**
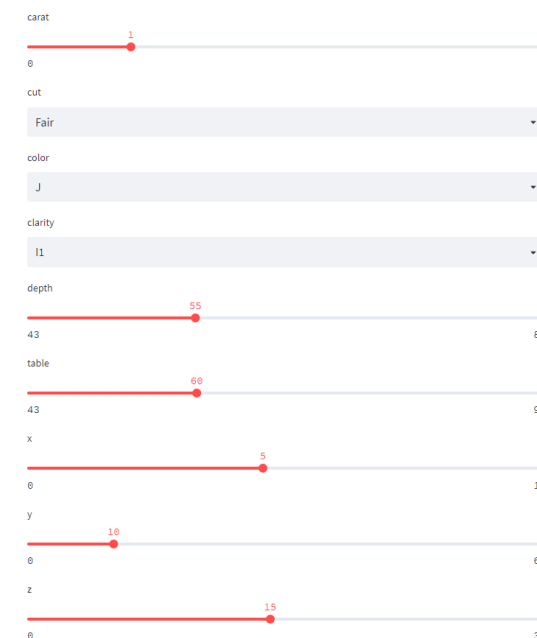
**The evaluation results of the overall model are good**

- **The error ratio** in the model with the overall data price is quite sufficient and we can consider it reasonable **(the error is acceptable with a ratio of ~1/10).**
- For price range **diamonds with a low majority price** ($300 to $2400), the model **accuracy might be a little bit poor** (because the model error is in the range of ~ $300 or less,
- but for diamonds priced above $2400 (especially for prices > $10K), **the overall error of the model is perfectly acceptable.**



Price Histogram (Chart 1)

*Min$ 326*
*Median $2.400*
*Max $18.800*

**Streamlit** **heroku**

# THANK YOU

"The smallest act of kindness is worth more than the grandest intention."
**Oscar Wilde (Playwright & Novelist)**

**06/Sept/2022**