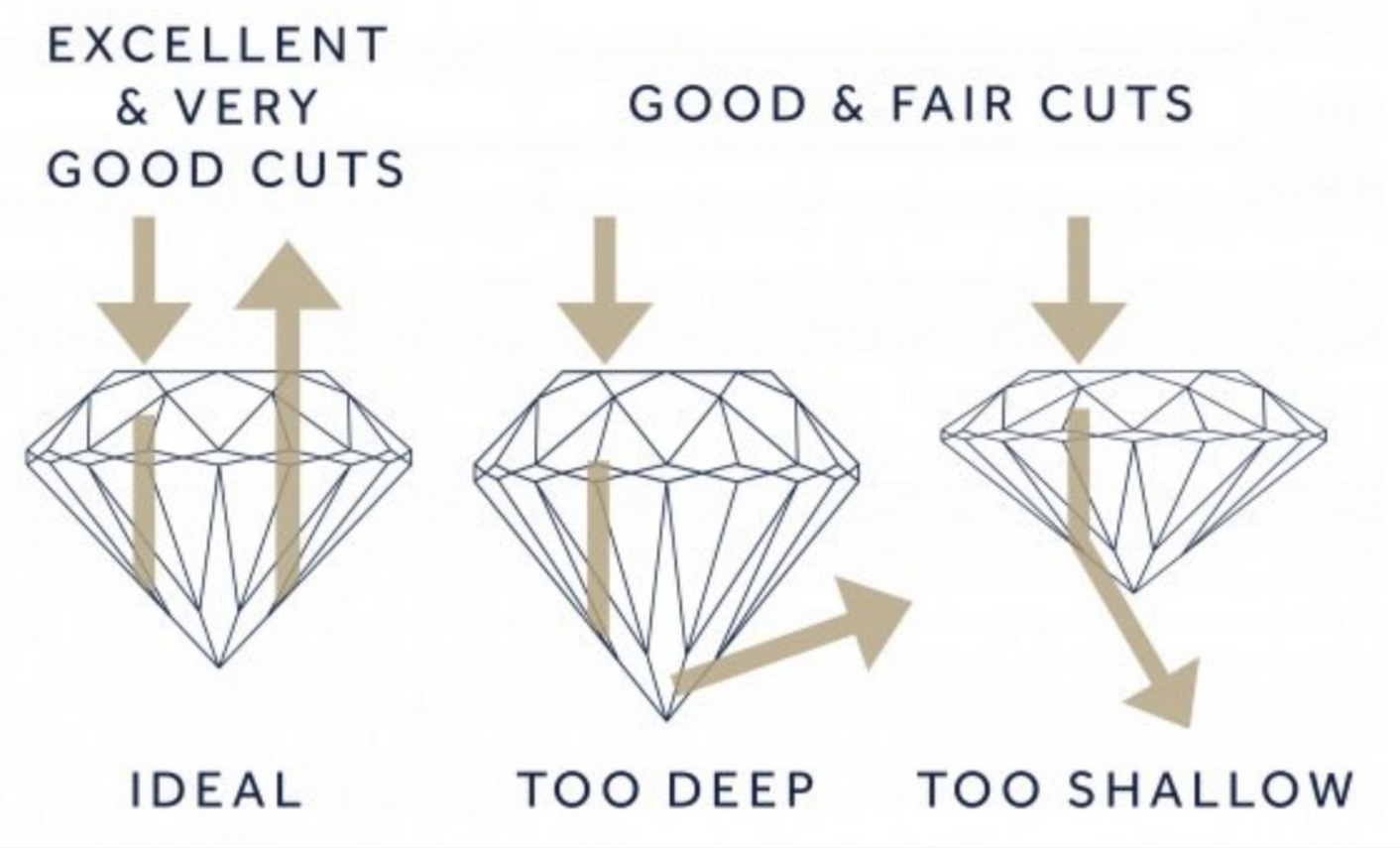
Diamonds Classification

Reef Alturki

Cut Quality

Diamond Cut is how well a diamond is cut and polished.



Problem Description

55% of all Round diamonds receive an excellent cut grade from the GIA.
About 25-30% of these excellent cut diamonds are not recommended.

THE DIAMOND PRO

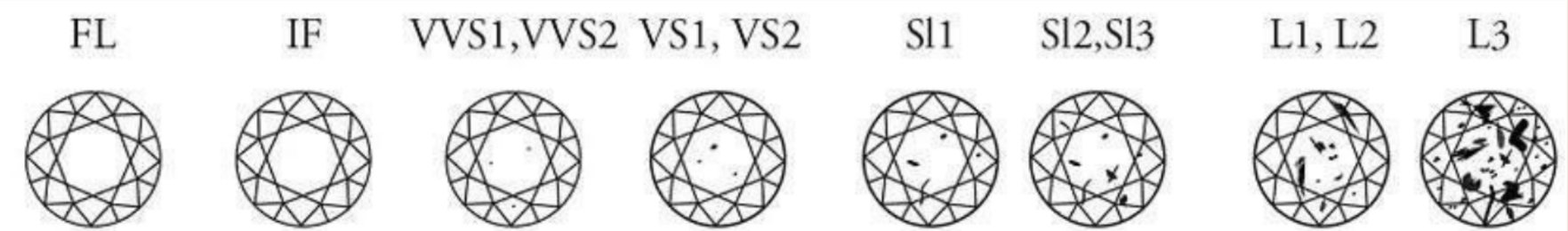
Claimed that their consultants review thousands of Excellent cut diamonds and find bad specs (depth, table and angles).

Dataset

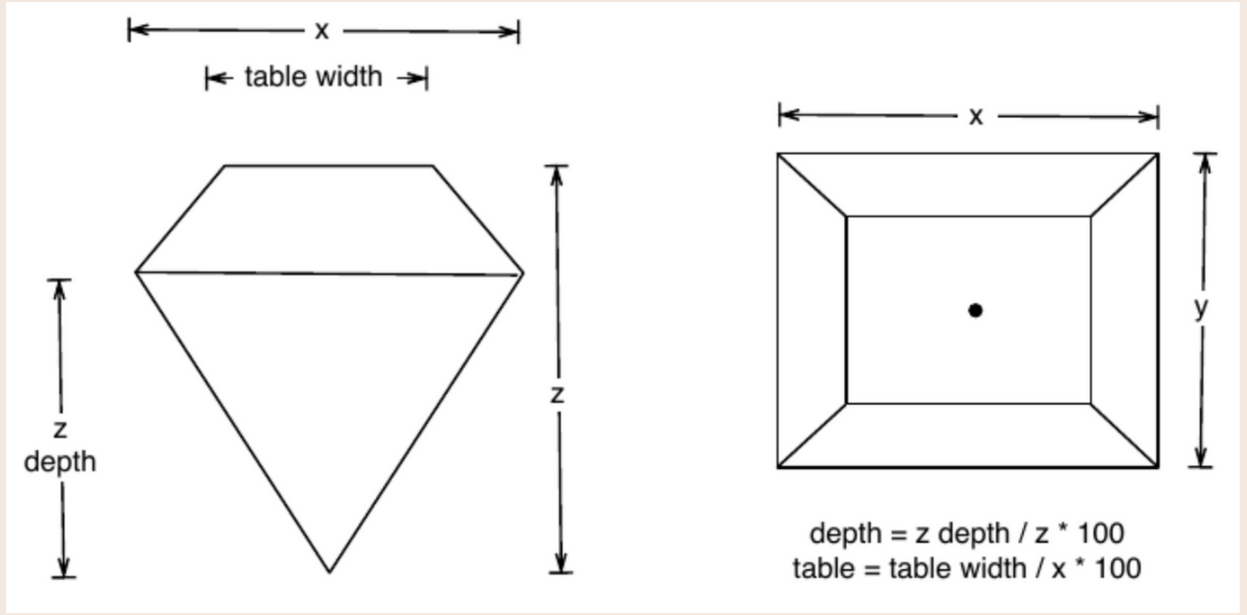
	carat	cut	color	clarity	depth	table	price	x	y	z
0	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
1	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
2	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
3	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63

Data contains 10 columns and 53940 observations.

Clarity



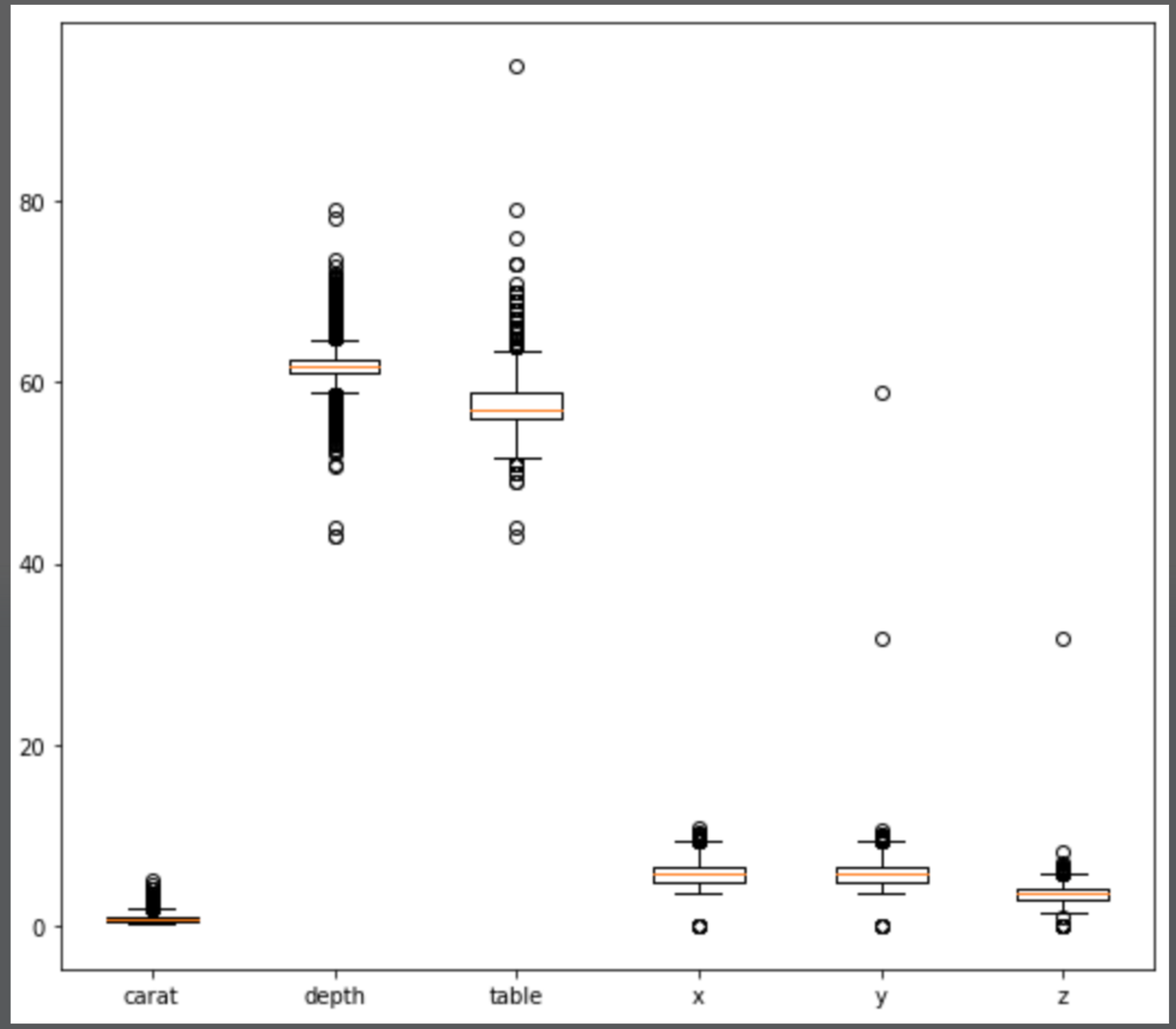
Dimensions



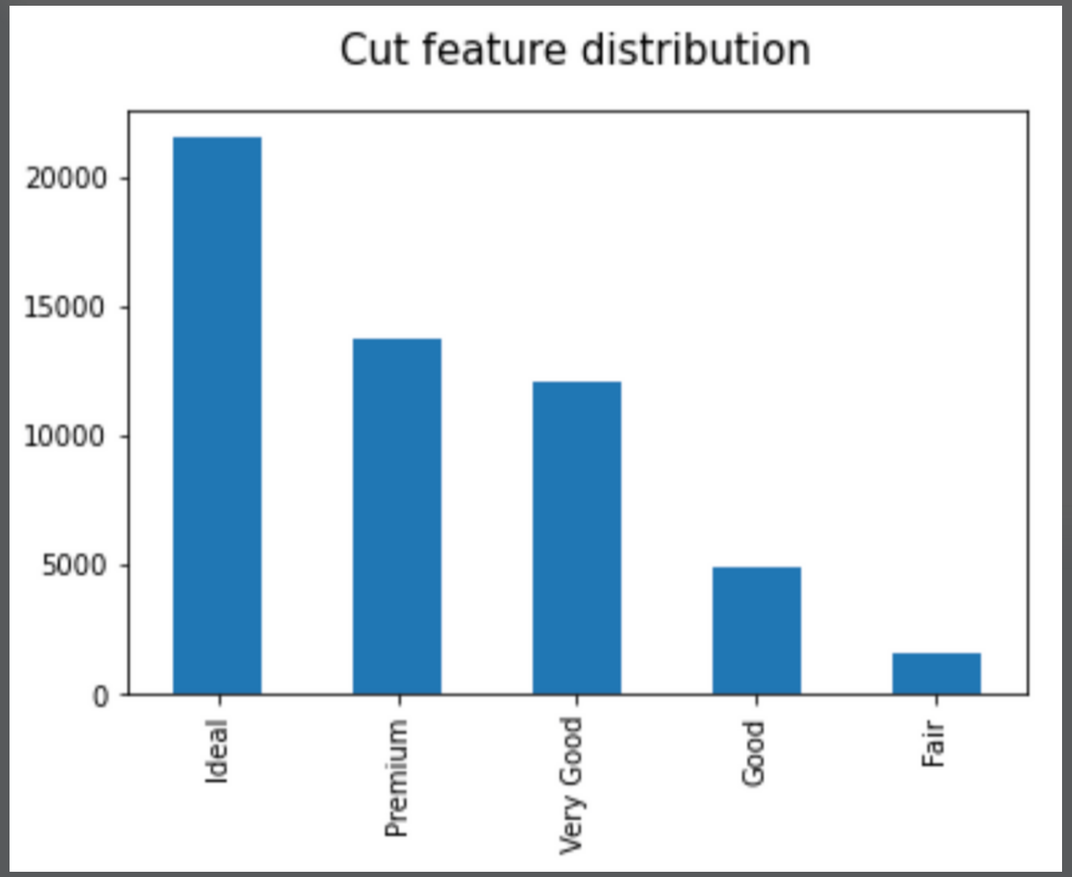
Color

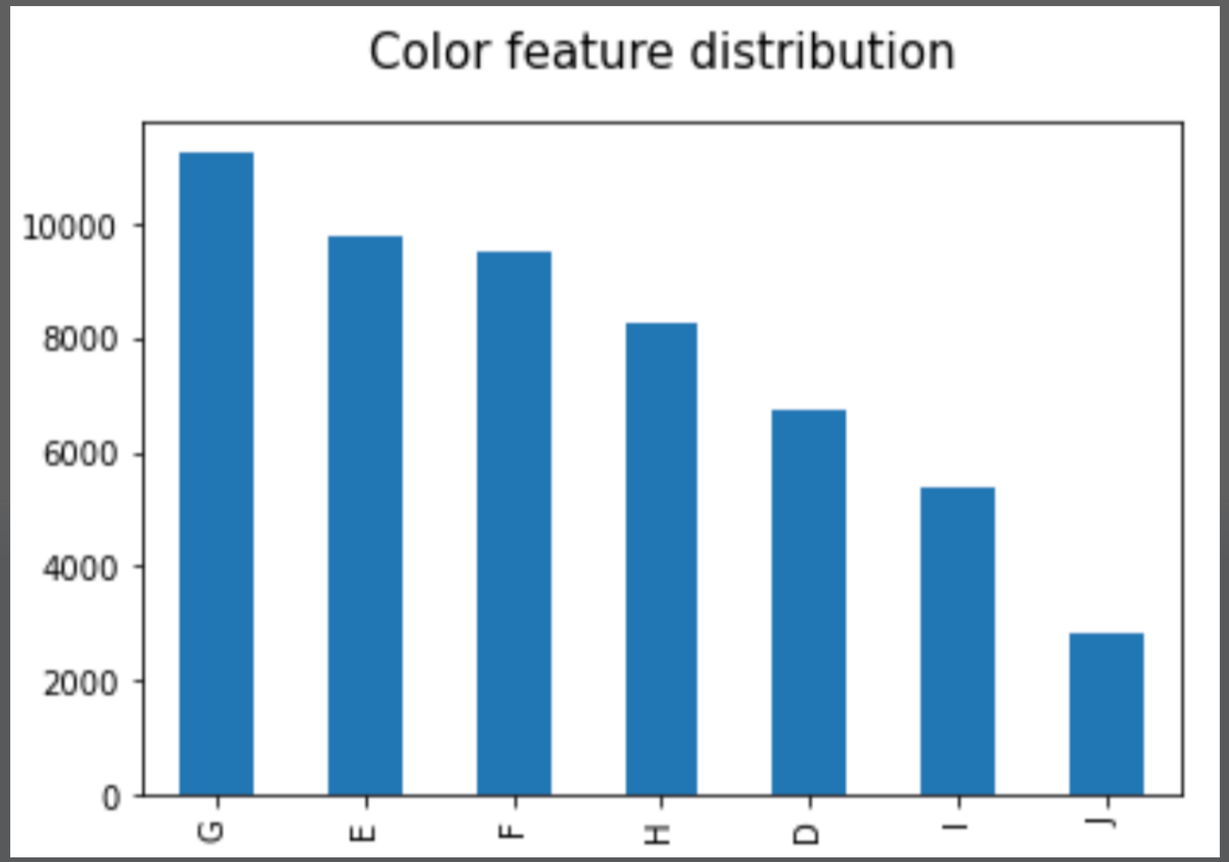
Data Cleaning

- Checking data types.
- Dropping duplicated rows.
- Detecting and removing outliers.

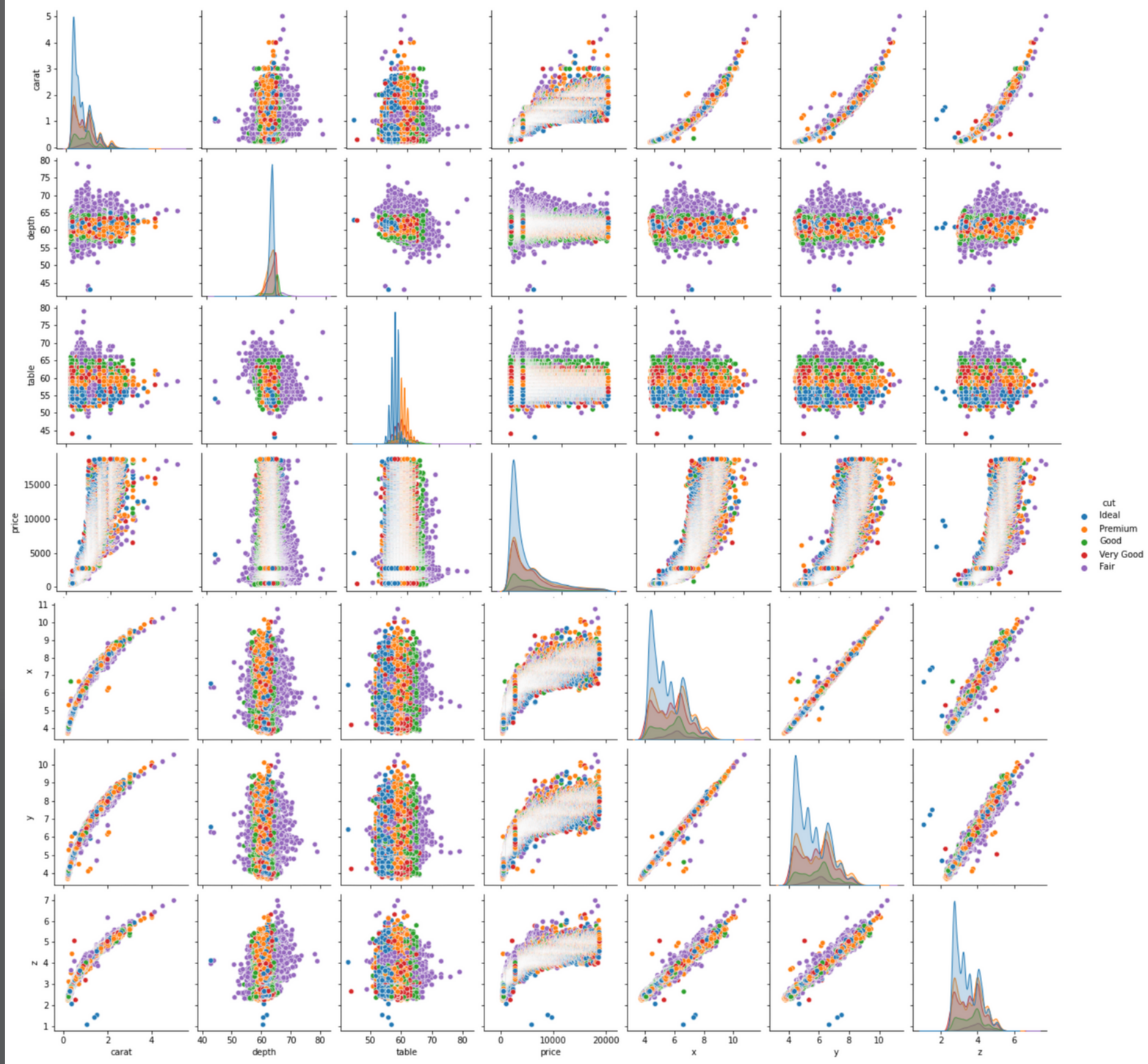


EDA


Cut feature distribution
20000
15000
10000
5000
0
Ideal
Premium
Very Good
Good
Fair

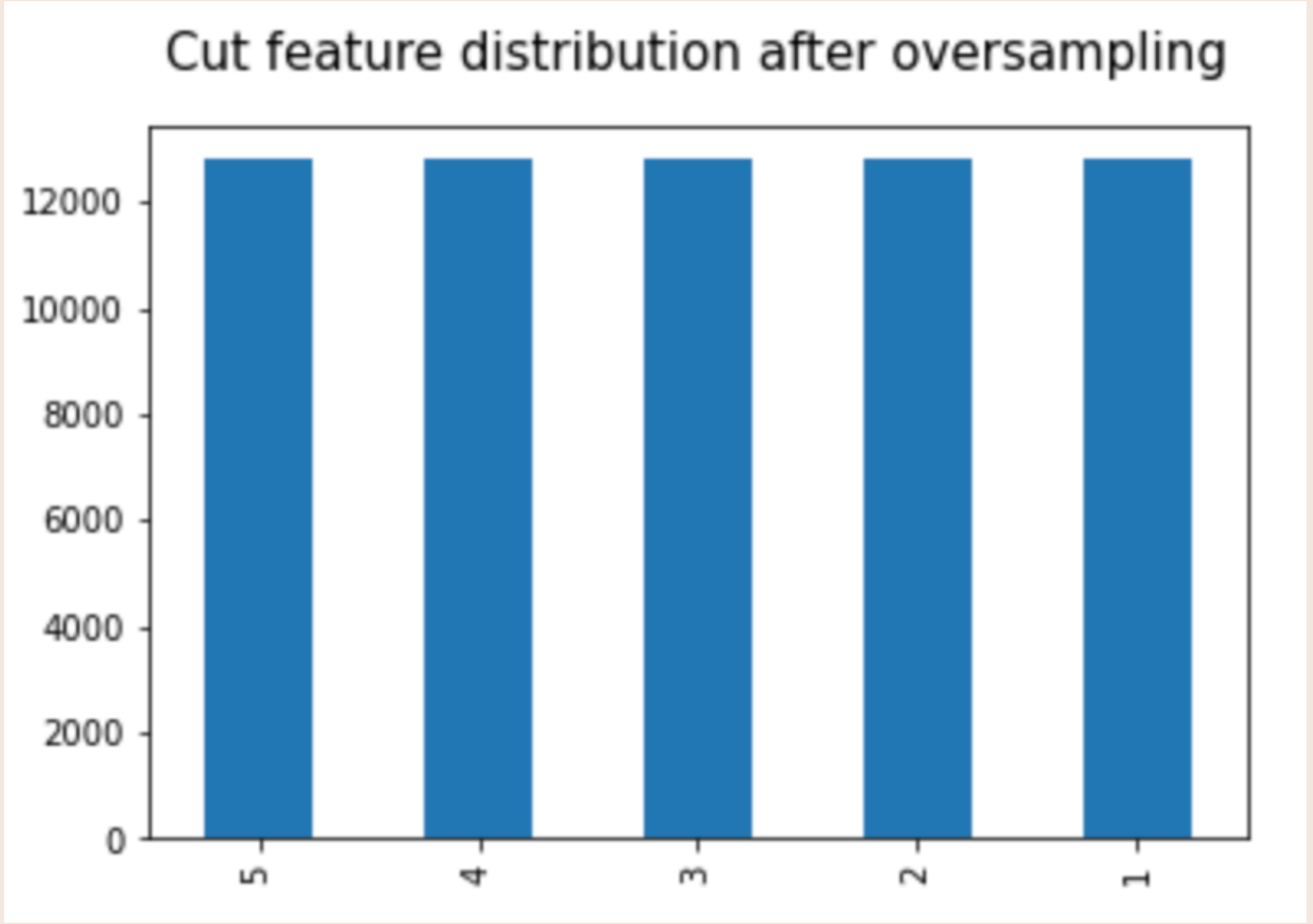

Color feature distribution
10000
8000
6000
4000
2000
0
G
E
F
H
D
I
J

Pairplot


carat
depth
table
price
x
y
z
cut
Ideal
Premium
Good
Very Good
Fair

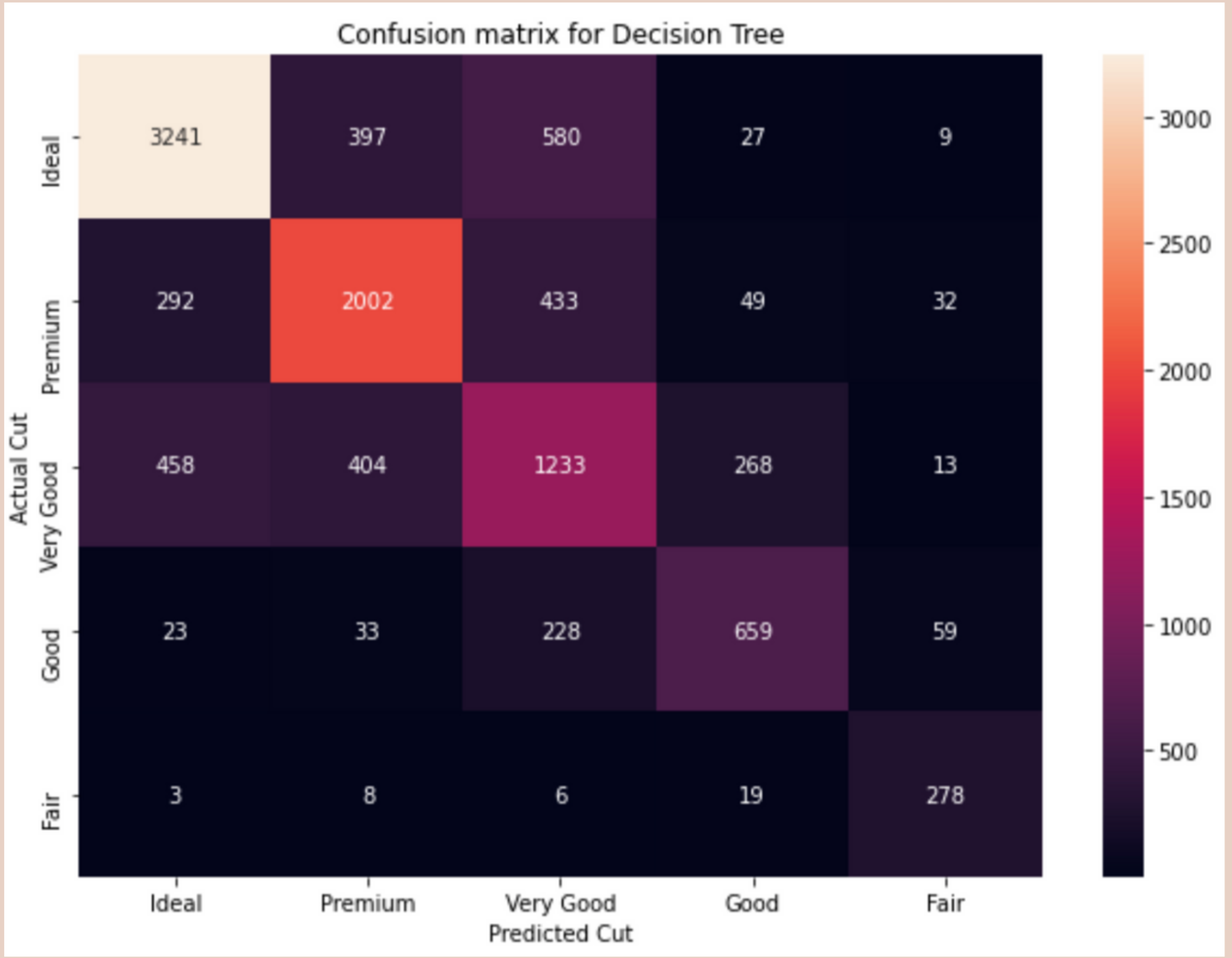
Imbalanced Classes

- SMOTE technique.
- Random Oversampling
- ADASYN technique.



Confusion matrix

The most appropriate classification metric is F1 score, as it balances between the precision and recall.



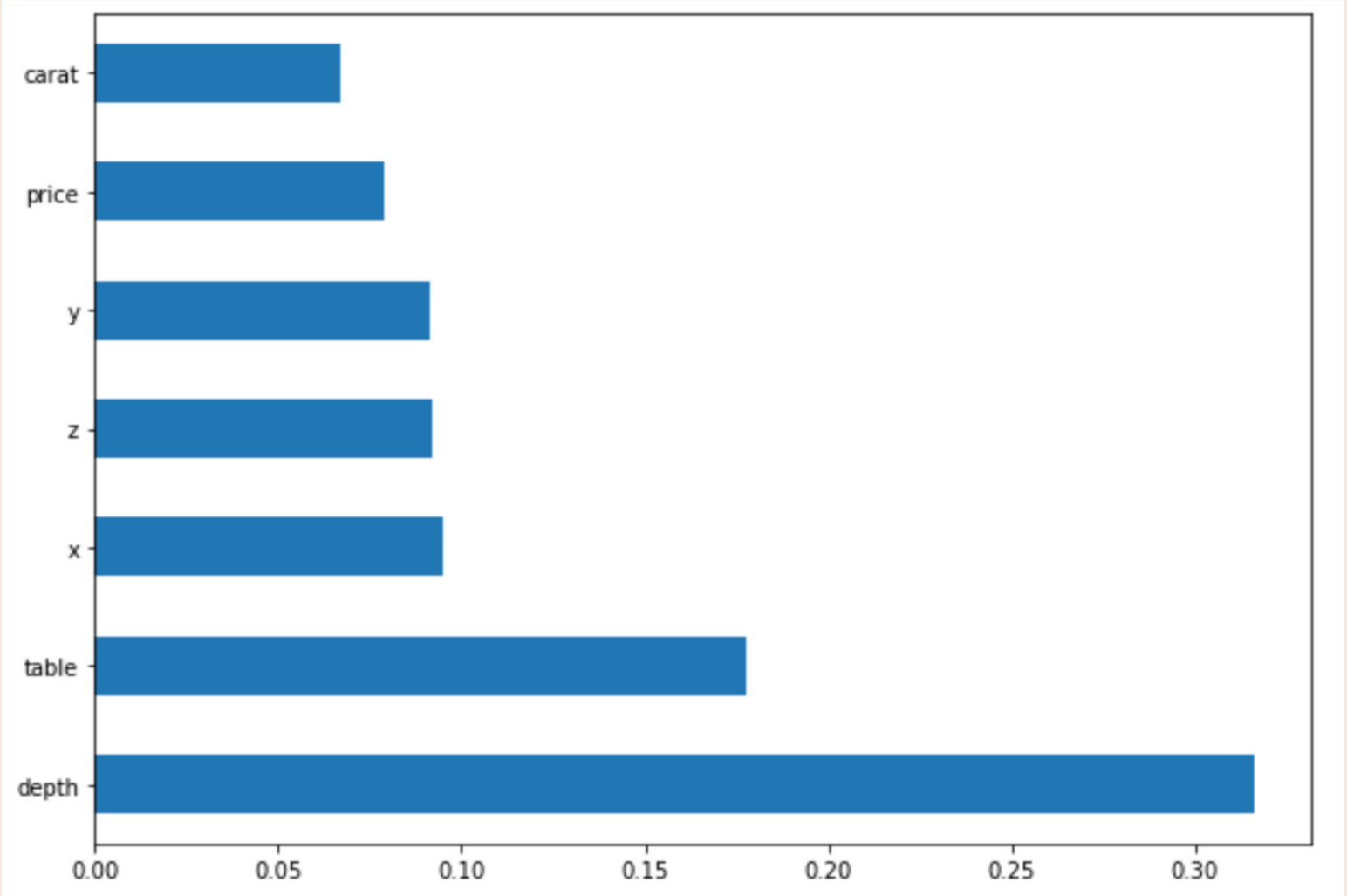
Modeling

Baseline Model	F1 score
• KNeighborsClassifier	0.636
Model Optimization	
• Logistic regression	0.614
• SGD Classifier	0.624
• Decision Tree	0.726
• Random Forest	0.753
• SVM Classifier	0.736

Final model

The Random Forest model achieved f1 score of 0.76 on the test set.

The plot shows feature importances for the Random Forest model.



Thank you