– Topic and tasks for the project

– Present your static data visualization and discuss how you design the data visualization

– Present your interactive data visualization and how to use it to address your tasks. Include a demo to show how to use your interactive data.

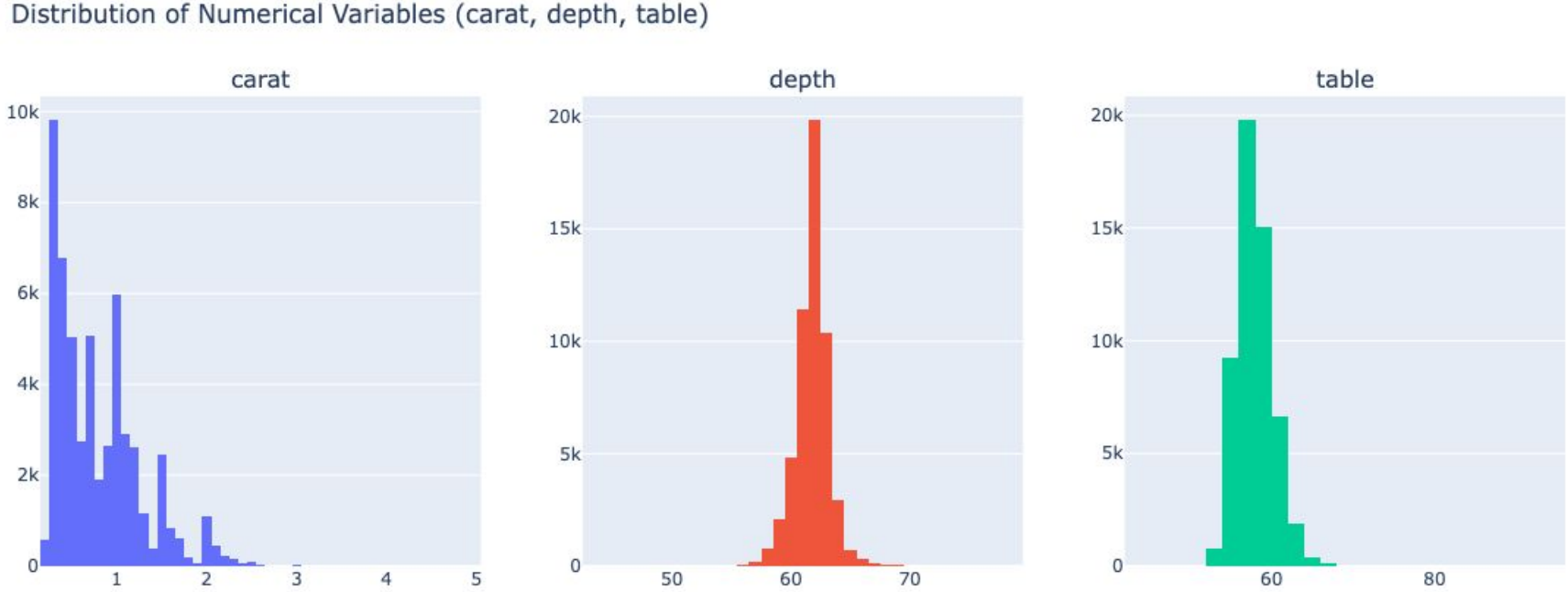
– Conclude what you have learned from the visualization

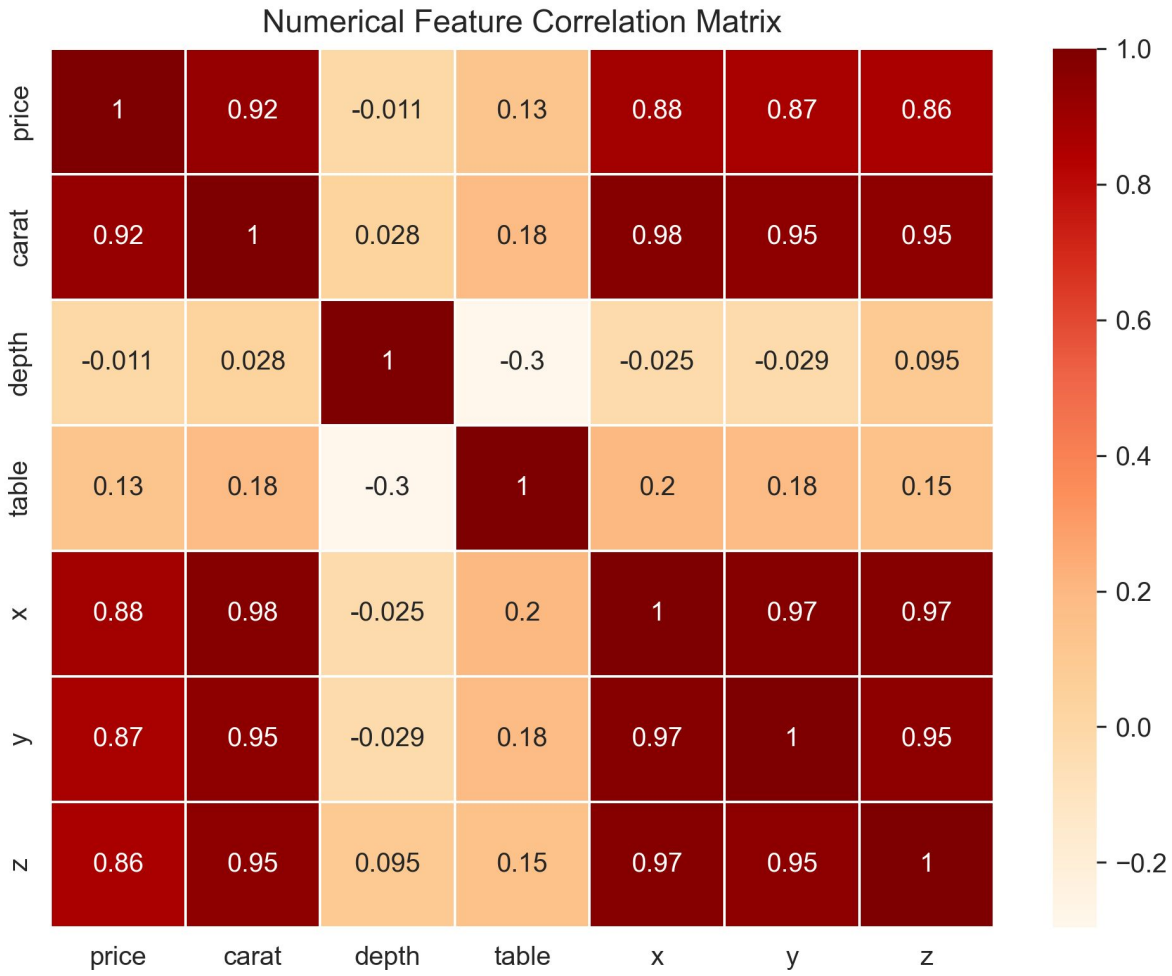
- 3.27 meeting

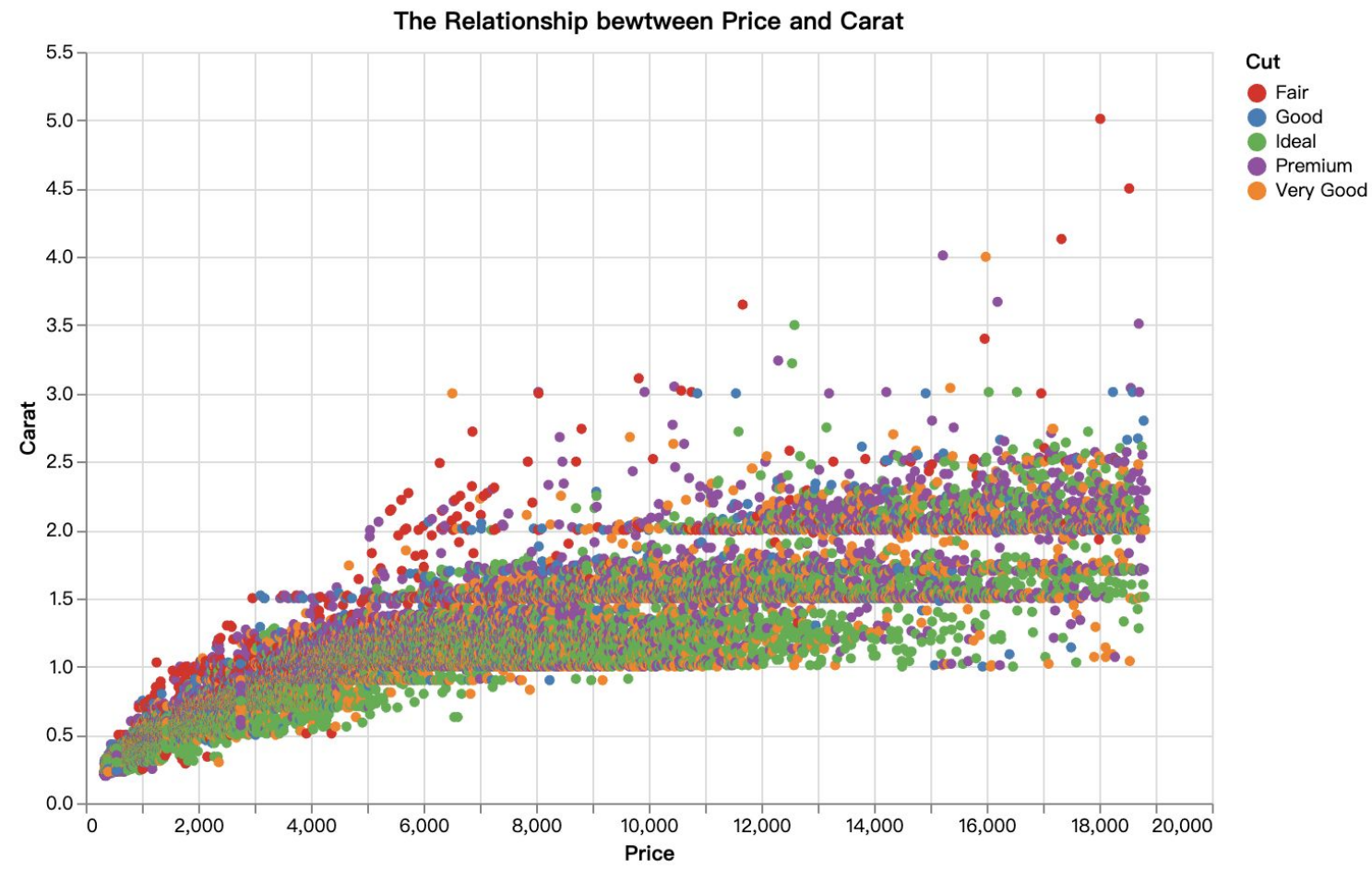
visual

1. bar plot - categorical data present
2. Density plot - numerical data present
3. Pairwise relationship
4. Correlation heatmap
5. box plot - categorical & price
6. Feature importance - numerical & price
7. D3 price & xyz

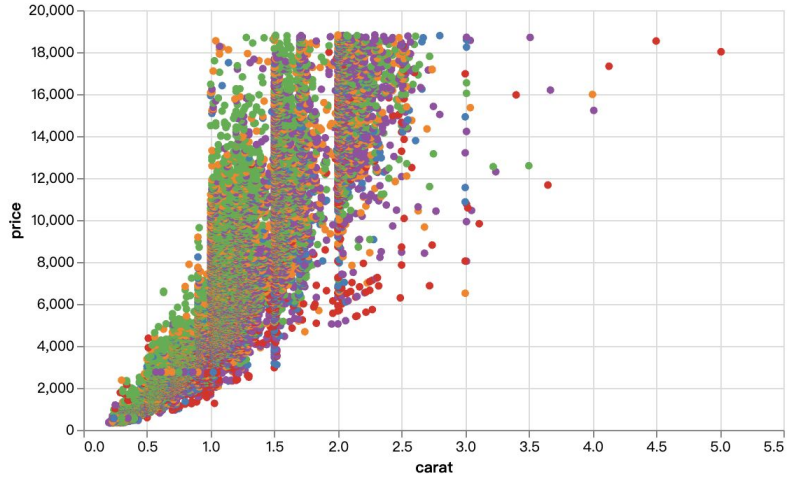

Count Distribution of Cut, Color, and Clarity
Count by Cut
20k
15k
10k
5k
0
Ideal
Premium
Very Good
Good
Fair
Count by Color
10k
8k
6k
4k
2k
0
G
E
F
H
D
I
J
Count by Clarity
12k
10k
8k
6k
4k
2k
0
SI1
VS2
SI2
VS1
VVS2
VVS1
IF
I1

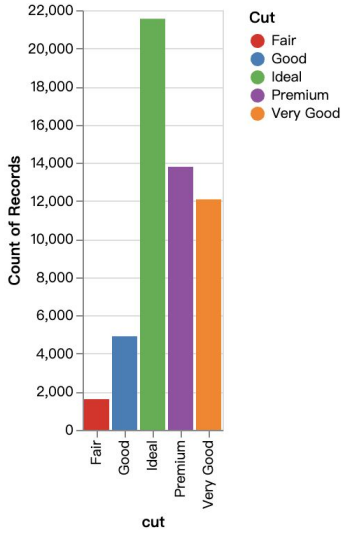

Distribution of Numerical Variables (carat, depth, table)
carat
depth
table
10k
8k
6k
4k
2k
0
1
2
3
4
5
20k
15k
10k
5k
0
50
60
70
20k
15k
10k
5k
0
60
80


Numerical Feature Correlation Matrix
price
carat
depth
table
x
y
z
1
0.92
-0.011
0.13
0.88
0.87
0.86
0.92
1
0.028
0.18
0.98
0.95
0.95
-0.011
0.028
1
-0.3
-0.025
-0.029
0.095
0.13
0.18
-0.3
1
0.2
0.18
0.15
0.88
0.98
-0.025
0.2
1
0.97
0.97
0.87
0.95
-0.029
0.18
0.97
1
0.95
0.86
0.95
0.095
0.15
0.97
0.95
1
price
carat
depth
table
x
y
z
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

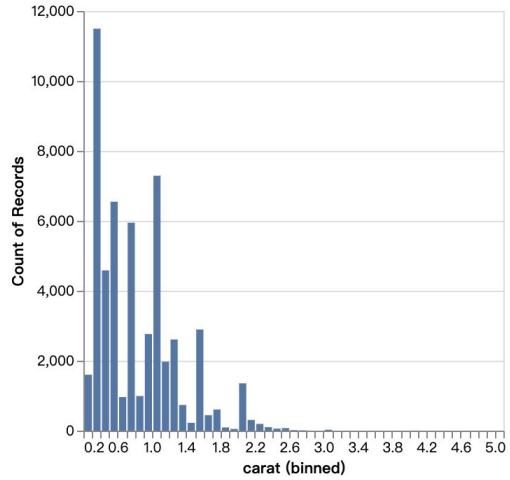


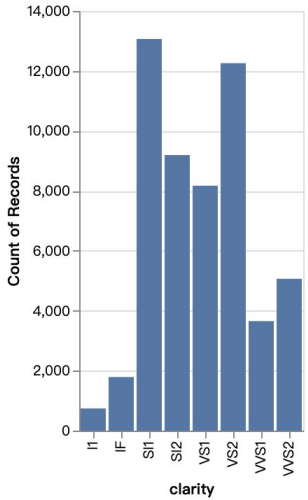
Cut: ◉ All ○ Fair ○ Good ○ Ideal ○ Premium ○ Very Good

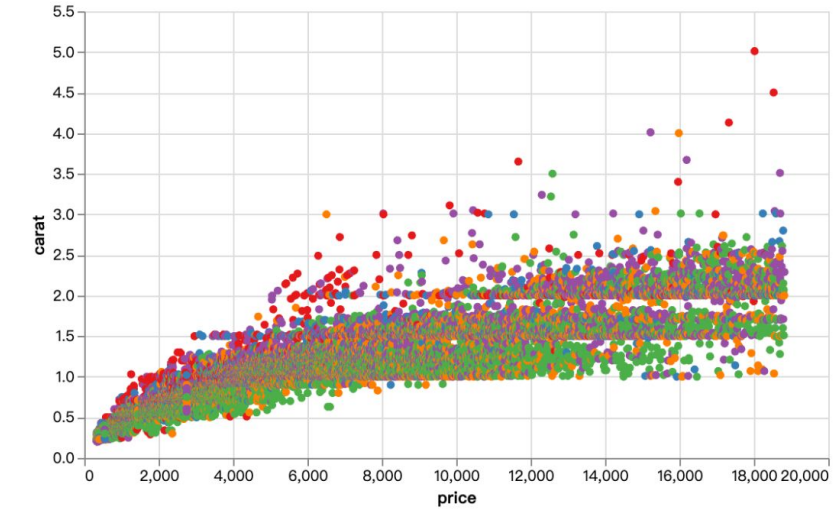

price
carat
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5

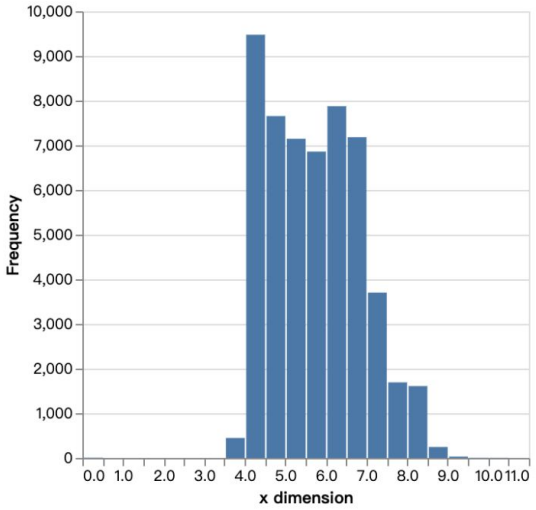

Count of Records
cut
Fair
Good
Ideal
Premium
Very Good
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000

Cut
Fair
Good
Ideal
Premium
Very Good

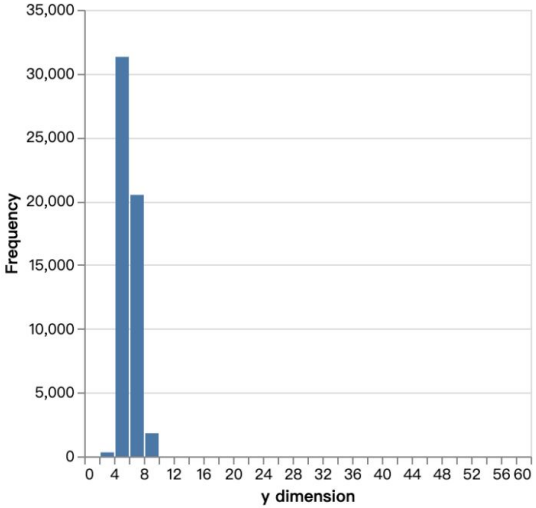

Count of Records
carat (binned)
0
2,000
4,000
6,000
8,000
10,000
12,000
0.2
0.6
1.0
1.4
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6
5.0

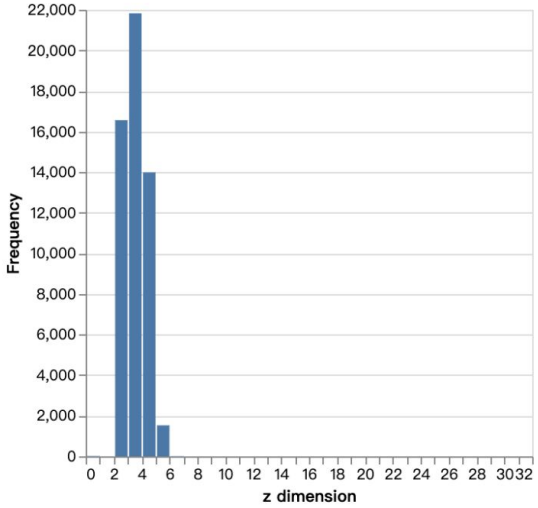

Count of Records
clarity
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
I1
IF
SI1
SI2
VS1
VS2
VVS1
VVS2

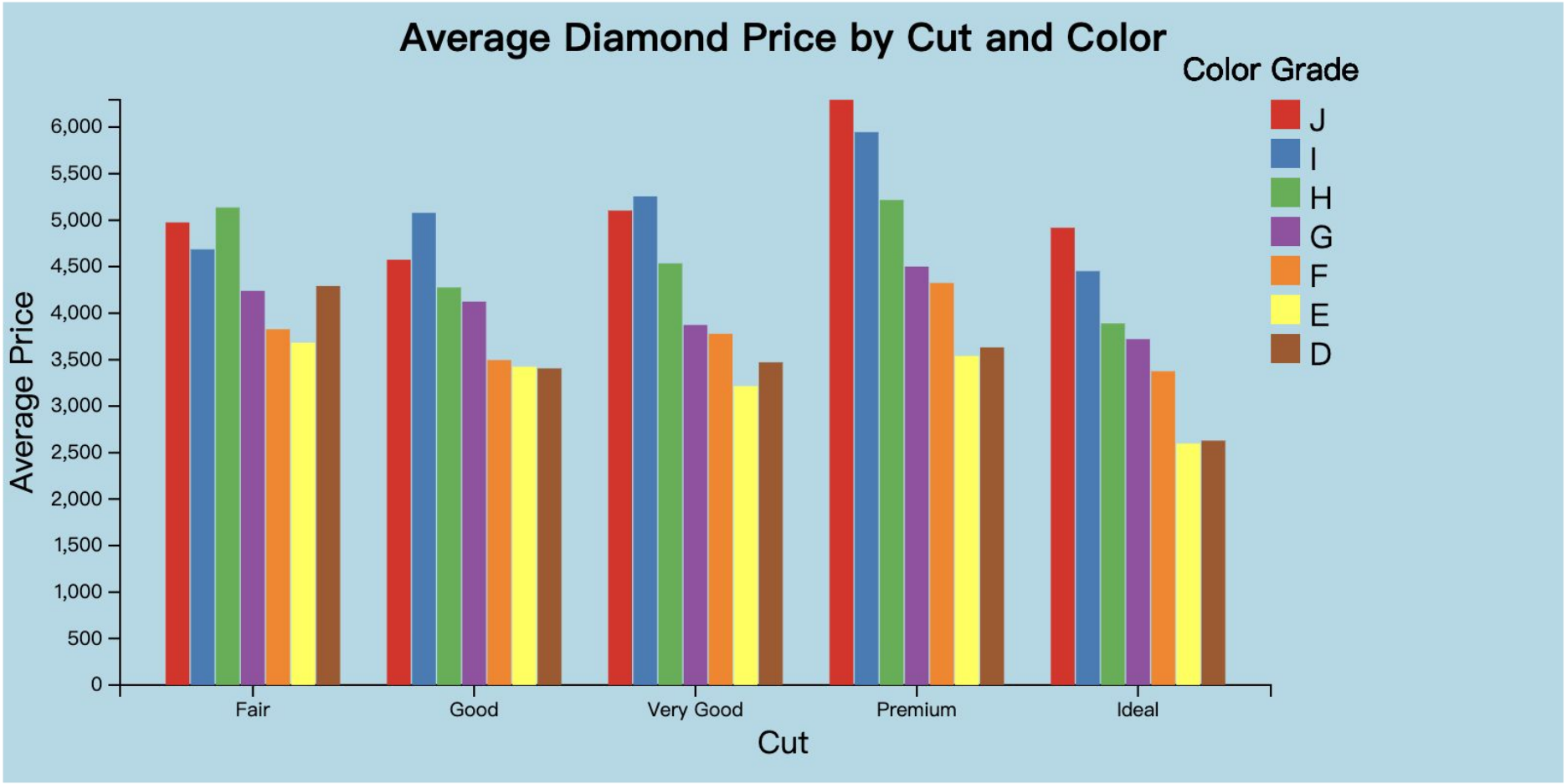

carat
price
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000

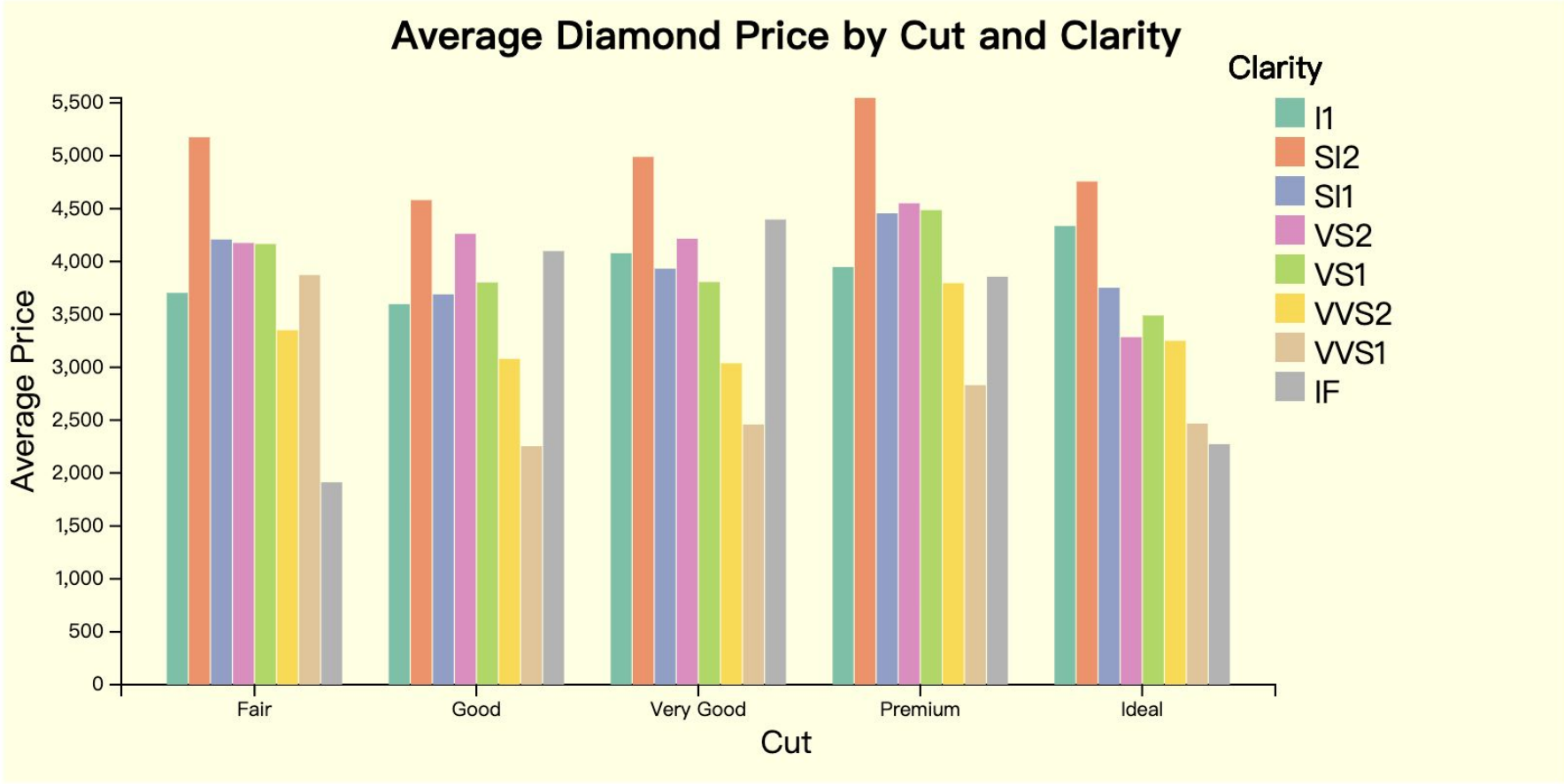

Frequency
x dimension
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0


Cut
Fair
Good
Ideal
Premium
Very Good

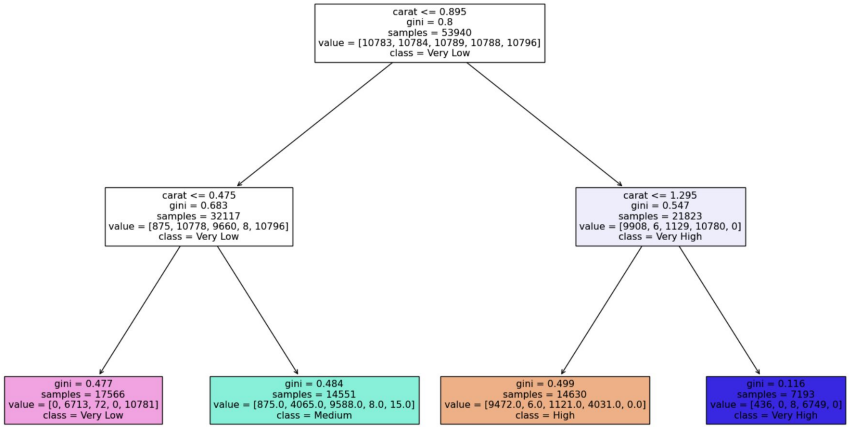

Frequency
y dimension
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
0
4
8
12
16
20
24
28
32
36
40
44
48
52
56
60

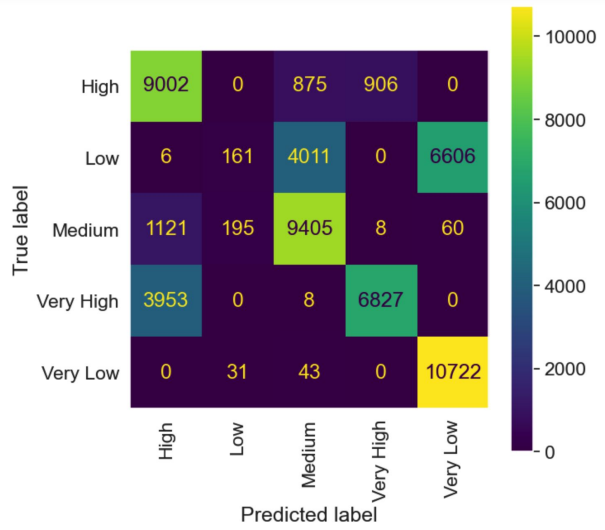

Frequency
z dimension
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32

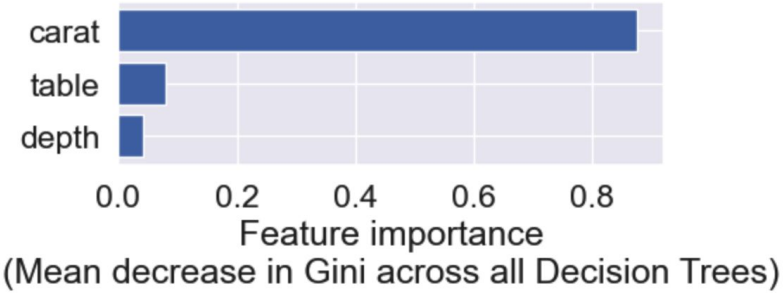

Average Diamond Price by Cut and Color
Color Grade
J
I
H
G
F
E
D
Average Price
6,000
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Fair
Good
Very Good
Premium
Ideal
Cut


Average Diamond Price by Cut and Clarity
Clarity
I1
SI2
SI1
VS2
VS1
VVS2
VVS1
IF
Average Price
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Fair
Good
Very Good
Premium
Ideal
Cut

Decision Tree Classifier and feature importance for numerical variables







Analysis of numerical variables