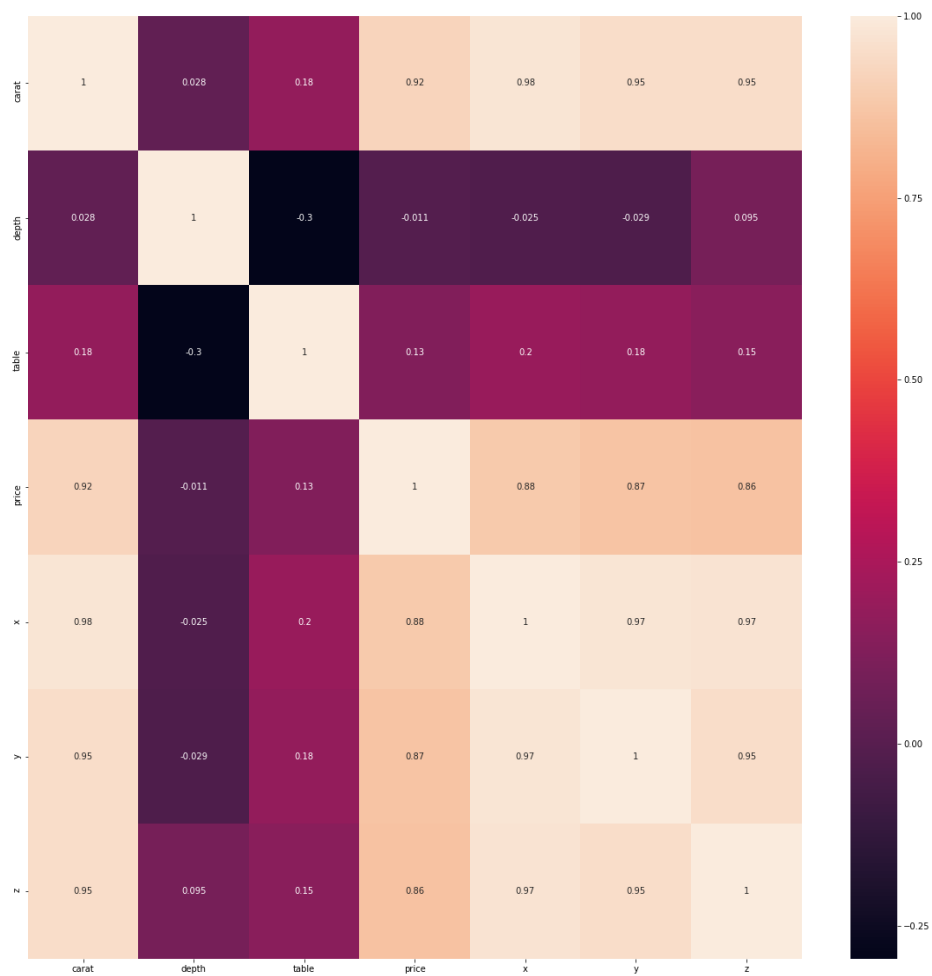# DIAMONDS ANALYSIS

- **Total Data: m = 53940, n = 11**
- **Price: Target Variable[continuous variable]**
- **Carat weight : continuous variable**
- **Cut,color,clarity : Discrete variable**
- **X,Y,Z : Length in MM**
- **depth total depth percentage = z / mean(x, y) = 2 * z / (x + y)**
- **table%: The width of the diamond's table expressed as a percentage of its average diameter**

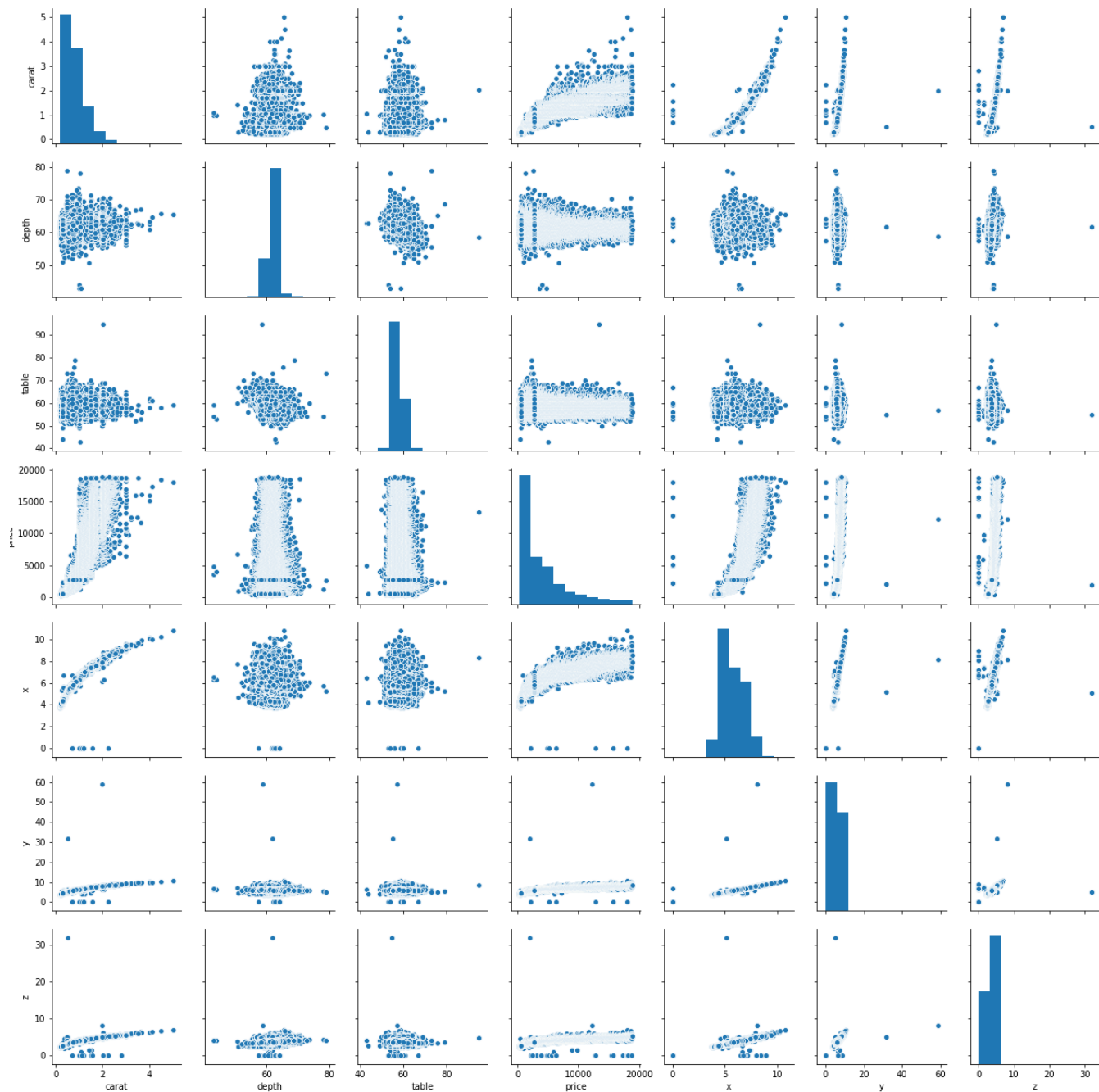'Unnamed: 0', 'carat', 'cut', 'color', 'clarity', 'depth', 'table', 'price', 'x','y', 'z'

Removed Column

'carat', 'cut', 'color', 'clarity', 'depth', 'table', 'price', 'x','y', 'z'

1. x,y, and z have a very strong relation with price but surprisingly depth (which comes from x,y, and z) doesn't has a significant relation with price.
2. Carat has a strong relation with price
3. Table doesn't have a significant relation with price or any other variable as well (We can try dropping that when making the model)

	carat	depth	table	price	x	y	z
carat	1	0.028	0.18	0.92	0.98	0.95	0.95
depth	0.028	1	-0.3	-0.011	-0.025	-0.029	0.095
table	0.18	-0.3	1	0.13	0.2	0.18	0.15
price	0.92	-0.011	0.13	1	0.88	0.87	0.86
x	0.98	-0.025	0.2	0.88	1	0.97	0.97
y	0.95	-0.029	0.18	0.87	0.97	1	0.95
z	0.95	0.095	0.15	0.86	0.97	0.95	1

|	carat	depth	table	price	x	y	z
count	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000	53940.000000
mean	0.797940	61.749405	57.457184	3932.799722	5.731157	5.734526	3.538734
std	0.474011	1.432621	2.234491	3989.439738	1.121761	1.142135	0.705699
min	0.200000	43.000000	43.000000	326.000000	0.000000	0.000000	0.000000
25%	0.400000	61.000000	56.000000	950.000000	4.710000	4.720000	2.910000
50%	0.700000	61.800000	57.000000	2401.000000	5.700000	5.710000	3.530000
75%	1.040000	62.500000	59.000000	5324.250000	6.540000	6.540000	4.040000
max	5.010000	79.000000	95.000000	18823.000000	10.740000	58.900000	31.800000

<u>X, Y,Z</u> values cannot be 0[Length ,width, Height cannot be zero]

X – 8

Y- 7

Z – 20

```
M = 53920, N = 10
```
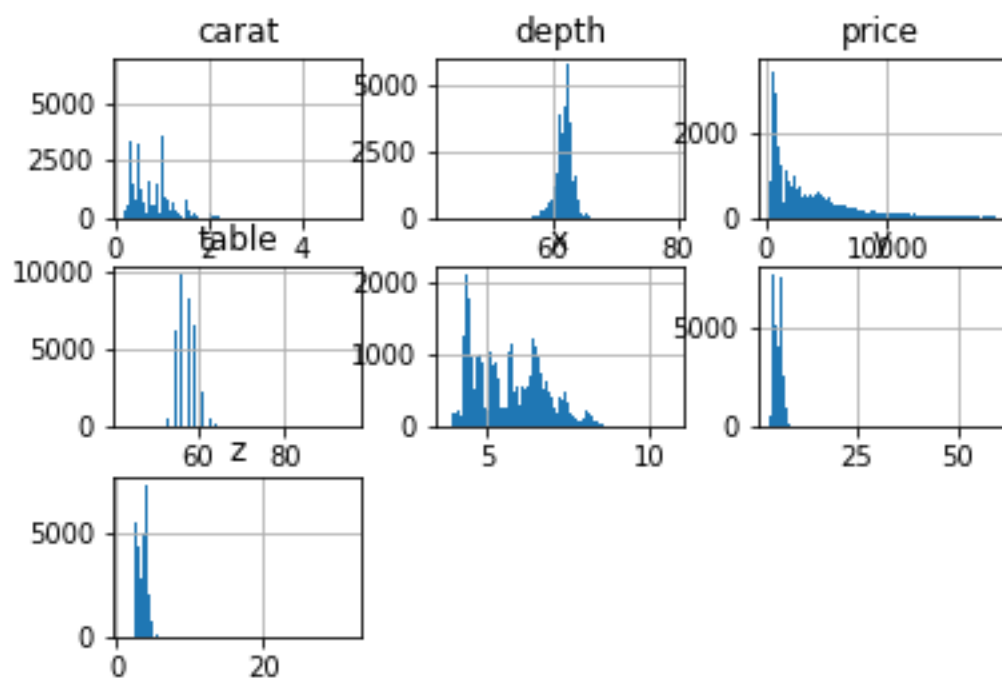
Catogarical:
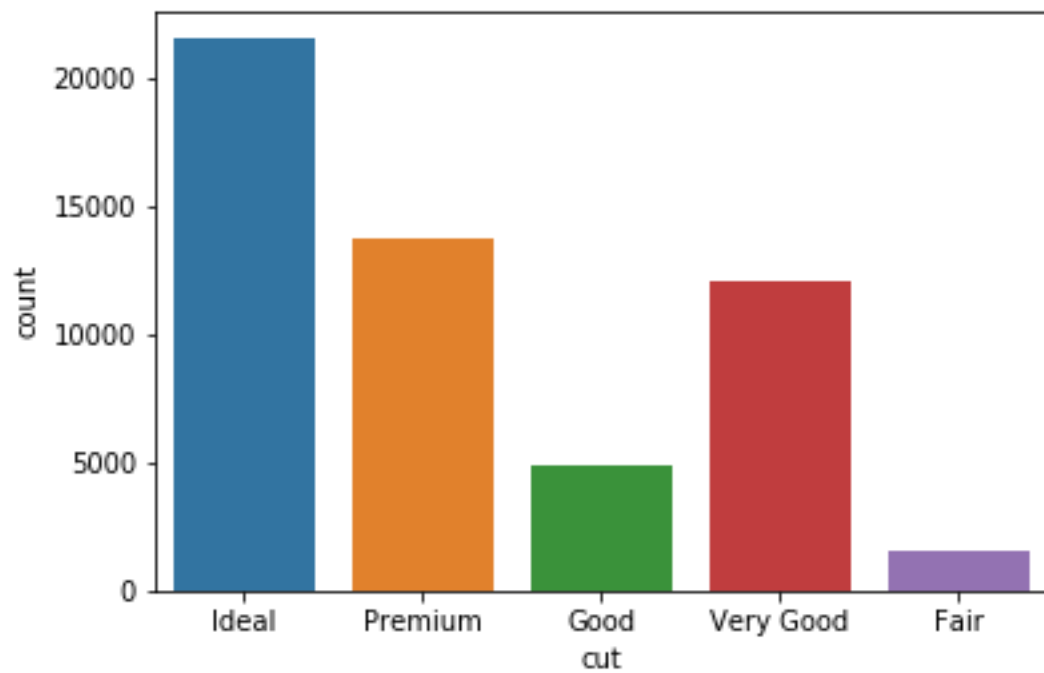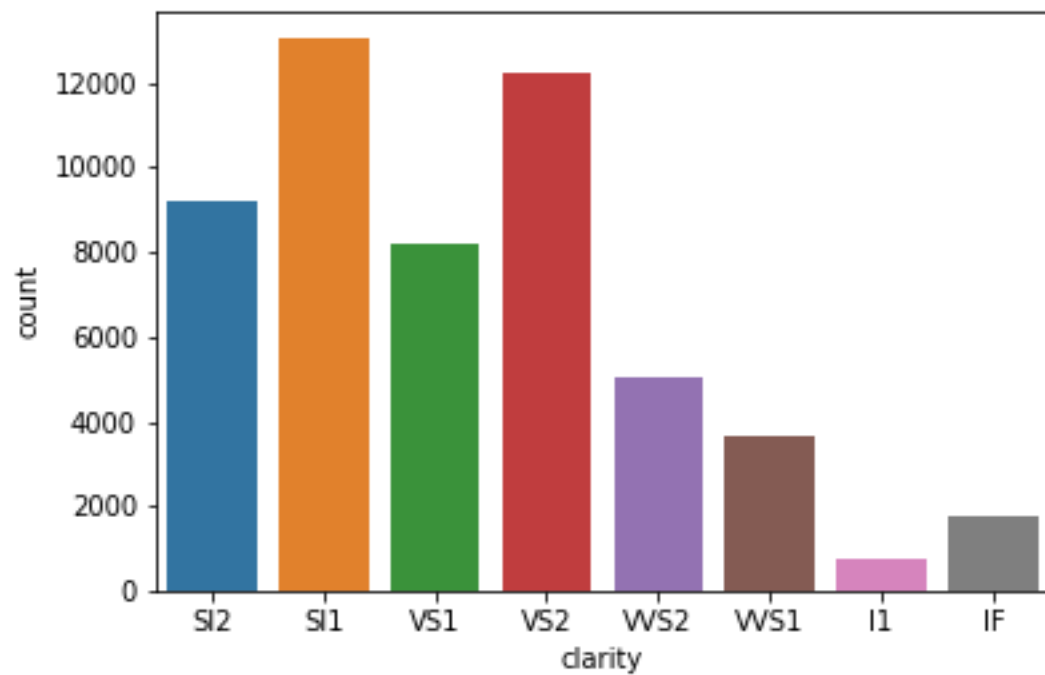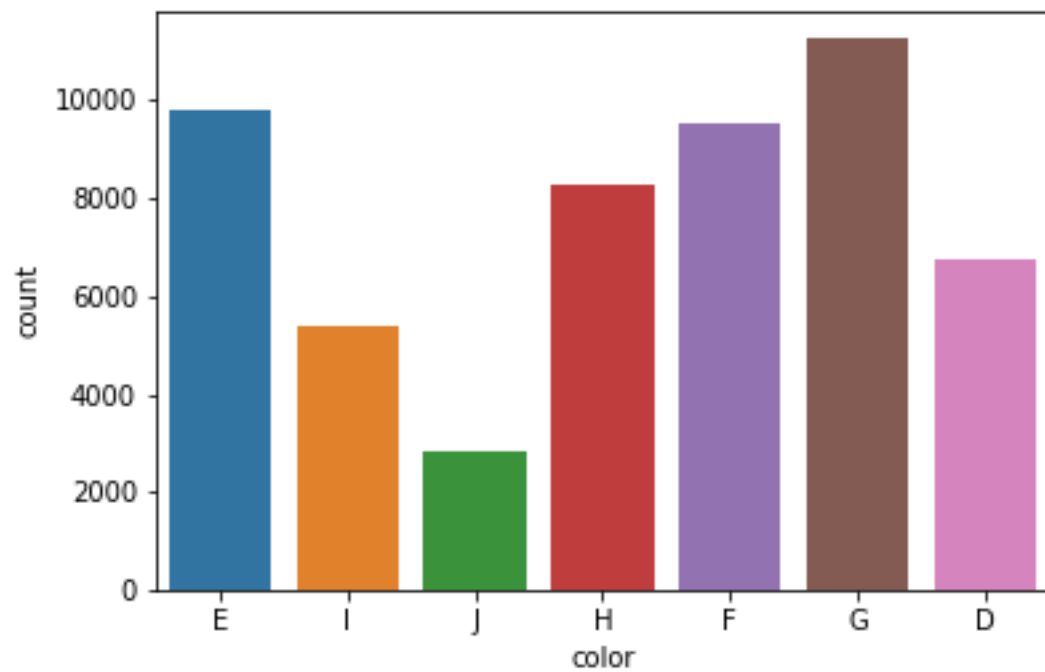
Created dummies variable

Scaled continuous variable

Train test split.