

```

... Dataset Shape: (53940, 10)
Columns: ['carat', 'cut', 'color', 'clarity', 'depth', 'table', 'price', 'x', 'y', 'z']

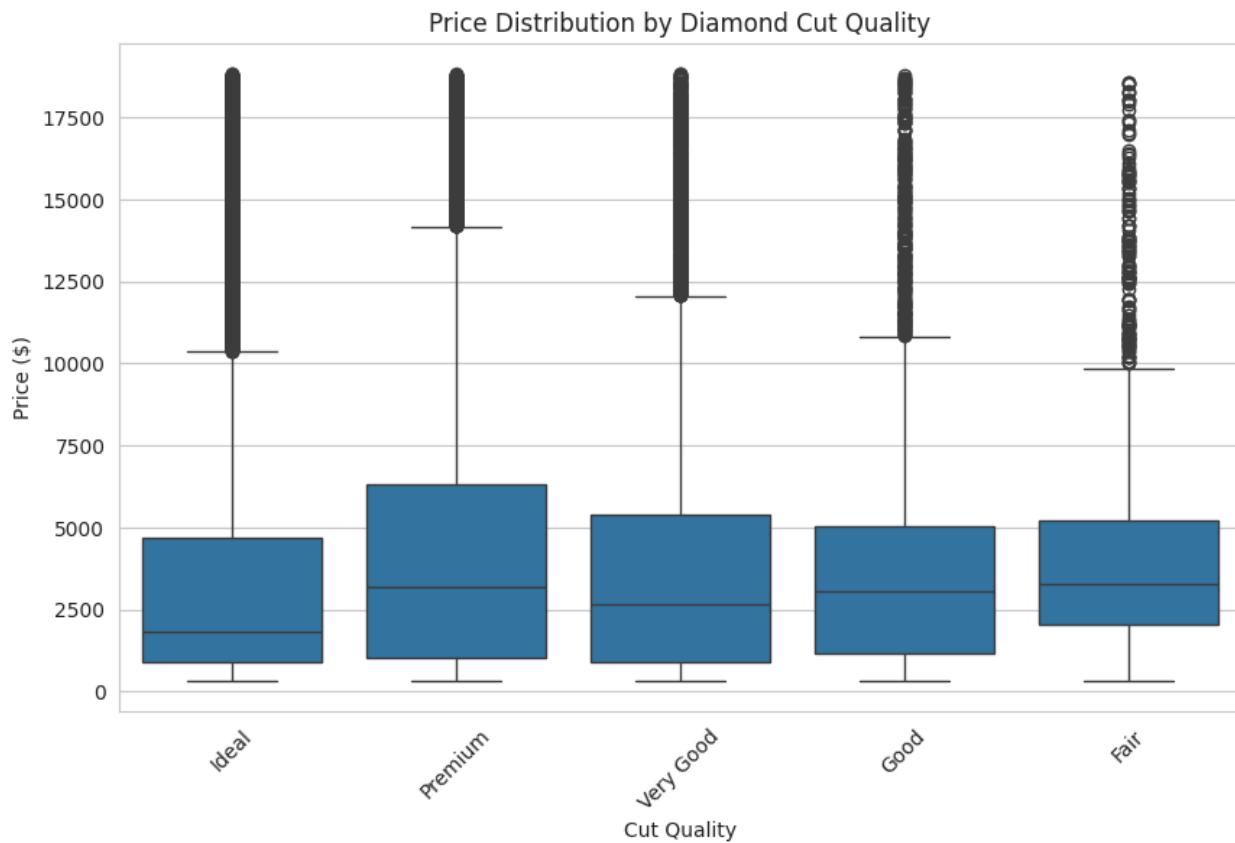
First 5 rows:
   carat      cut color clarity depth  table price     x     y     z
0  0.23    Ideal    E     SI2   61.5   55.0    326  3.95  3.98  2.43
1  0.21  Premium    E     SI1   59.8   61.0    326  3.89  3.84  2.31
2  0.23     Good    E     VS1   56.9   65.0    327  4.05  4.07  2.31
3  0.29  Premium    I     VS2   62.4   58.0    334  4.20  4.23  2.63
4  0.31     Good    J     SI2   63.3   58.0    335  4.34  4.35  2.75

```

The output shows the first 5 rows with all columns (carat, cut, color, clarity, depth, table, price, x, y, z).



This scatter plot shows a strong positive correlation between carat weight and price, with prices increasing exponentially rather than linearly as carat size increases.



The box plot reveals that while 'Ideal' and 'Premium' cuts generally command higher prices, there is significant overlap across all cut categories, indicating other factors (clarity, color) heavily influence price.

```
BEFORE - Missing Values Check:
carat      0
cut        0
color      0
clarity    0
depth      0
table      0
price      0
x          0
y          0
z          0
dtype: int64
Zero dimensions (invalid data): x      8
y      7
z     20
dtype: int64
AFTER - Shape changed from (53940, 10) to (53920, 10)
Missing/Invalid values handled by removing rows with zero dimensions
```