First 5 rows of raw data:

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

0 1 0.23 Ideal E SI2 61.5 55.0 3.95 3.98 2.43 326

1 2 0.21 Premium E SI1 59.8 61.0 3.89 3.84 2.31 326

2 3 0.23 Good E VS1 56.9 65.0 4.05 4.07 2.31 327

3 4 0.29 Premium I VS2 62.4 58.0 4.20 4.23 2.63 334

4 5 0.31 Good J SI2 63.3 58.0 4.34 4.35 2.75 335

Last 5 rows of raw data:

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

53935 53936 0.72 Ideal D SI1 60.8 57.0 5.75 5.76 3.50 2757

53936 53937 0.72 Good D SI1 63.1 55.0 5.69 5.75 3.61 2757

53937 53938 0.70 Very Good D SI1 62.8 60.0 5.66 5.68 3.56 2757

53938 53939 0.86 Premium H SI2 61.0 58.0 6.15 6.12 3.74 2757

53939 53940 0.75 Ideal D SI2 62.2 55.0 5.83 5.87 3.64 2757

Shape of raw data:

(53940, 11)

Any missing data:

False

Data Statistical Description:

Unnamed: 0 carat depth table x y z price

count 53940.000000 53940.000000 53940.000000 53940.000000 53940.000000 53940.000000 53940.000000 53940.000000

mean 26970.500000 0.797940 61.749405 57.457184 5.731157 5.734526 3.538734 3932.799722

std 15571.281097 0.474011 1.432621 2.234491 1.121761 1.142135 0.705699 3989.439738

min 1.000000 0.200000 43.000000 43.000000 0.000000 0.000000 0.000000 326.000000

25% 13485.750000 0.400000 61.000000 56.000000 4.710000 4.720000 2.910000 950.000000

50% 26970.500000 0.700000 61.800000 57.000000 5.700000 5.710000 3.530000 2401.000000

75% 40455.250000 1.040000 62.500000 59.000000 6.540000 6.540000 4.040000 5324.250000

max 53940.000000 5.010000 79.000000 95.000000 10.740000 58.900000 31.800000 18823.000000

Data Types Information:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 53940 entries, 0 to 53939

Data columns (total 11 columns):

Unnamed: 0 53940 non-null int64

carat 53940 non-null float64

cut 53940 non-null object

color 53940 non-null object

clarity 53940 non-null object

depth 53940 non-null float64

table 53940 non-null float64

x 53940 non-null float64

y 53940 non-null float64

z 53940 non-null float64

price 53940 non-null int64

dtypes: float64(6), int64(2), object(3)

memory usage: 4.5+ MB

None

Unnamed: 0 carat depth table x y z price

count 53917.000000 53917.000000 53917.000000 53917.000000 53917.000000 53917.000000 53917.000000 53917.000000

mean 26971.599904 0.797687 61.749565 57.456939 5.731605 5.733428 3.539409 3930.910474

std 15571.129462 0.473777 1.432318 2.234069 1.119402 1.111272 0.691620 3987.215003

min 1.000000 0.200000 43.000000 43.000000 3.730000 3.680000 1.070000 326.000000

25% 13487.000000 0.400000 61.000000 56.000000 4.710000 4.720000 2.910000 949.000000

50% 26973.000000 0.700000 61.800000 57.000000 5.700000 5.710000 3.530000 2401.000000

75% 40456.000000 1.040000 62.500000 59.000000 6.540000 6.540000 4.040000 5323.000000

max 53940.000000 5.010000 79.000000 95.000000 10.740000 10.540000 6.980000 18823.000000

carat cut color clarity depth table x y z price

carat 1.000000 -0.133886 -0.291834 -0.352420 0.025661 0.180381 0.977725 0.976766 0.976124 0.920550

cut -0.133886 1.000000 0.018778 0.193935 -0.221843 -0.430870 -0.124279 -0.124109 -0.151100 -0.051714

color -0.291834 0.018778 1.000000 -0.023335 -0.039772 -0.029004 -0.271167 -0.270898 -0.274503 -0.171801

clarity -0.352420 0.193935 -0.023335 1.000000 -0.069140 -0.160340 -0.372530 -0.367028 -0.376398 -0.144360

depth 0.025661 -0.221843 -0.039772 -0.069140 1.000000 -0.296166 -0.027419 -0.030469 0.094480 -0.013172

table 0.180381 -0.430870 -0.029004 -0.160340 -0.296166 1.000000 0.193819 0.187474 0.153607 0.127041

x 0.977725 -0.124279 -0.271167 -0.372530 -0.027419 0.193819 1.000000 0.998650 0.990865 0.886688

y 0.976766 -0.124109 -0.270898 -0.367028 -0.030469 0.187474 0.998650 1.000000 0.990517 0.888242

z 0.976124 -0.151100 -0.274503 -0.376398 0.094480 0.153607 0.990865 0.990517 1.000000 0.881241

price 0.920550 -0.051714 -0.171801 -0.144360 -0.013172 0.127041 0.886688 0.888242 0.881241 1.000000

LinearRegression[default]

R2: 0.91

MSE: 1441048.40

RMSE: 1200.44

MAE: 803.56

Coefficients: [ 5109.55354319 148.16900859 549.18277036 817.83646258 75.18116752 -32.30489783 -1968.64351557 2414.21505948 -1453.05863271]

Intercept: 3924.2047375531997

Modelling Time: 00.015624 seconds

----------

KNeighborsRegressor[default]

R2: 0.97

MSE: 532509.66

RMSE: 729.73

MAE: 378.73

Modelling Time: 00.541984 seconds

----------

Ridge[default]

R2: 0.91

MSE: 1441231.18

RMSE: 1200.51

MAE: 803.63

Coefficients: [ 5105.21834433 147.86459132 549.09273112 818.12801528 72.87113728 -32.76510506 -1929.82856341 2361.85277609 -1435.05555303]

Intercept: 3924.2047375532006

Modelling Time: 00.031218 seconds

----------

DecisionTreeRegressor[default]

R2: 0.97

MSE: 548317.03

RMSE: 740.48

MAE: 360.77

Feature Importances: [0.6276657 0.00186726 0.03411314 0.06208856 0.00338148 0.00211516 0.00491409 0.25864788 0.00520674]

Modelling Time: 00.156279 seconds

----------

RandomForestRegressor[default]

R2: 0.98

MSE: 326569.13

RMSE: 571.46

MAE: 285.12

Feature Importances: [0.50903739 0.00178649 0.03236658 0.06304649 0.00318621 0.00227084 0.00486576 0.37789963 0.00554061]

Modelling Time: 00.993098 seconds

----------

GradientBoostingRegressor[default]

R2: 0.98

MSE: 387245.05

RMSE: 622.29

MAE: 344.58

Feature Importances: [0.14081038 0.01981612 0.22693461 0.1801286 0.02228763 0.00287875 0.06127623 0.19286811 0.15299957]

Modelling Time: 00.996107 seconds

----------

SGDRegressor[default]

R2: 0.91

MSE: 1464675.05

RMSE: 1210.24

MAE: 808.35

Coefficients: [4944.36510426 154.45908526 551.34302446 825.66112387 -62.18968314 -31.99650489 -442.22163211 95.30779833 -536.66976255]

Intercept: [3914.26248182]

Modelling Time: 00.041679 seconds

----------

SVR[default]

R2: 0.50

MSE: 8146634.59

RMSE: 2854.23

MAE: 1350.76

Modelling Time: 50.382575 seconds

----------

LinearSVR[default]

R2: 0.85

MSE: 2443995.18

RMSE: 1563.33

MAE: 850.75

Coefficients: [ 2.85775032e+03 8.76967989e+01 2.50867173e+02 4.42285383e+02 -8.16753189e+00 -2.36874847e+00 1.34411205e+02 1.83062891e+02 1.03454454e+02]

Intercept: [3449.98614164]

Modelling Time: 00.046809 seconds

----------

MLPRegressor[default]

R2: 0.96

MSE: 580956.19

RMSE: 762.20

MAE: 428.92

Loss: 289207.25318696426

Iteration Number: 200

Layer Number: 3

Output Number: 1

Modelling Time: 20.935457 seconds

----------

LinearRegression

R2: 0.91

MSE: 1441048.40

RMSE: 1200.44

MAE: 803.56

Coefficients: [ 5109.55354319 148.16900859 549.18277036 817.83646258 75.18116752 -32.30489783 -1968.64351557 2414.21505948 -1453.05863271]

Intercept: 3924.2047375531997

Modelling Time: 00.031246 seconds

----------

KNeighborsRegressor

R2: 0.97

MSE: 431043.42

RMSE: 656.54

MAE: 329.54

Modelling Time: 00.391984 seconds

----------

Ridge

R2: 0.91

MSE: 1441231.18

RMSE: 1200.51

MAE: 803.63

Coefficients: [5105.21834433 147.86459132 549.09273112 818.12801528 72.87113728 -32.76510506 -1929.82856341 2361.85277609 -1435.05555303]

Intercept: 3924.2047375532006

Modelling Time: 00.030781 seconds

----------

DecisionTreeRegressor

R2: 0.97

MSE: 559611.03

RMSE: 748.07

MAE: 362.08

Feature Importances: [0.62777271 0.00194498 0.03413543 0.0621006 0.0033454 0.00212296 0.0046969 0.25862829 0.00525275]

Modelling Time: 00.164174 seconds

----------

RandomForestRegressor

R2: 0.98

MSE: 308918.89

RMSE: 555.80

MAE: 279.54

Feature Importances: [0.59397 0.00191137 0.03305643 0.06328145 0.00309846 0.0022551 0.00493141 0.29187598 0.00561978]

Modelling Time: 00.588720 seconds

----------

GradientBoostingRegressor

R2: 0.98

MSE: 274042.59

RMSE: 523.49

MAE: 272.61

Feature Importances: [0.14533936 0.03461504 0.08952316 0.07339138 0.08083547 0.04425581 0.15788501 0.21826531 0.15588945]

Modelling Time: 20.278558 seconds

----------

SGDRegressor

R2: 0.90

MSE: 1644775.39

RMSE: 1282.49

MAE: 820.27

Coefficients: [5321.19019039 291.12251111 332.93655793 798.21985972 19.69893318 -11.06452181 -1967.98749718 1856.25897151 -1510.96212989]

Intercept: [3845.15409845]

Modelling Time: 00.046894 seconds

----------

SVR

R2: 0.89

MSE: 1734711.44

RMSE: 1317.08

MAE: 697.09

Modelling Time: 39.831736 seconds

----------

LinearSVR

R2: 0.91

MSE: 1439525.14

RMSE: 1199.80

MAE: 798.38

Coefficients: [ 5275.55413527 145.44229133 552.51050522 815.26632628 65.63375781 -38.4402649 -2036.88884265 2301.67241034 -1426.50853817]

Intercept: [3923.4208047]

Modelling Time: 02.744344 seconds

----------

MLPRegressor

R2: 0.98

MSE: 333576.81

RMSE: 577.56

MAE: 323.83

Modelling Time: 16.417136 seconds