PREDICTING JEWELRY PRICES

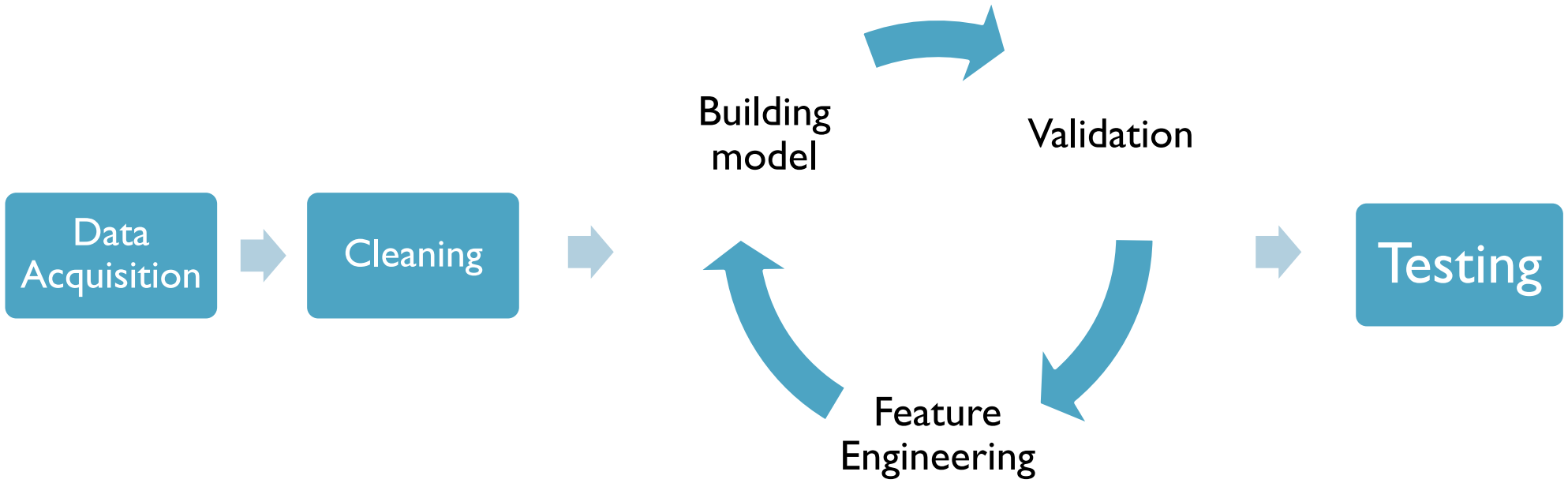
PROJECT 02

Can we automate the function of a jeweler?

OBJECTIVE & EXPECTED OUTCOME

- Deliver a machine Learning model that is able to predict or estimate the price of jewelries.

METHODOLOGY



DATA ACQUISITION

Tools:

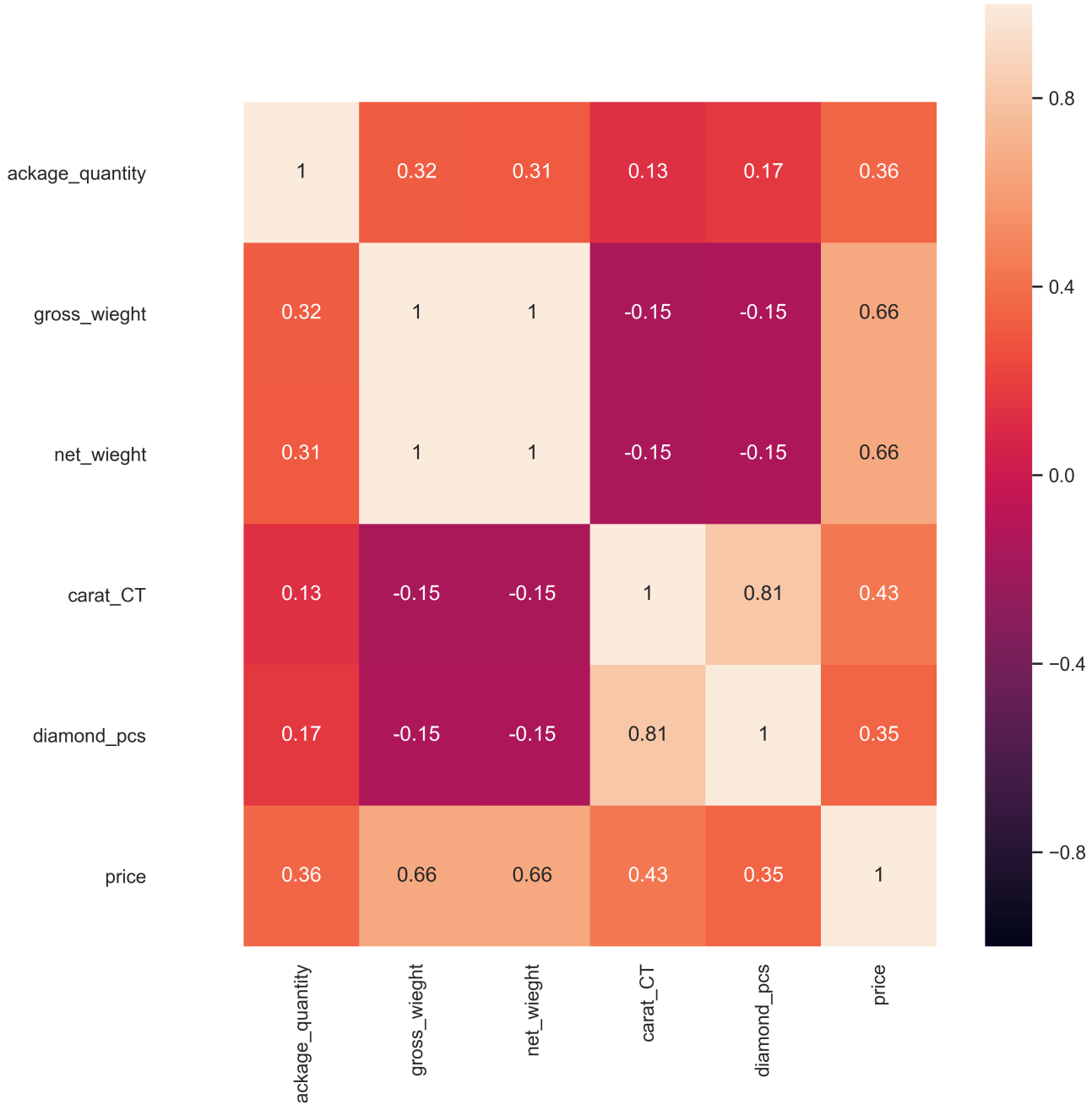
- Selenium
- Beautiful Soup

CLEANING

- Convert the data type into numeric.
- Handle missing values.
- Handling outliers (IQR).

item_package_quantity	gross_wieght	net_wieght	carat_CT	diamond_pcs
1.0	2.050	2.050	0.32	42.0
1.0	2.920	2.920	0.00	0.0
1.0	2.690	2.690	0.00	0.0
1.0	4.260	4.260	0.00	0.0

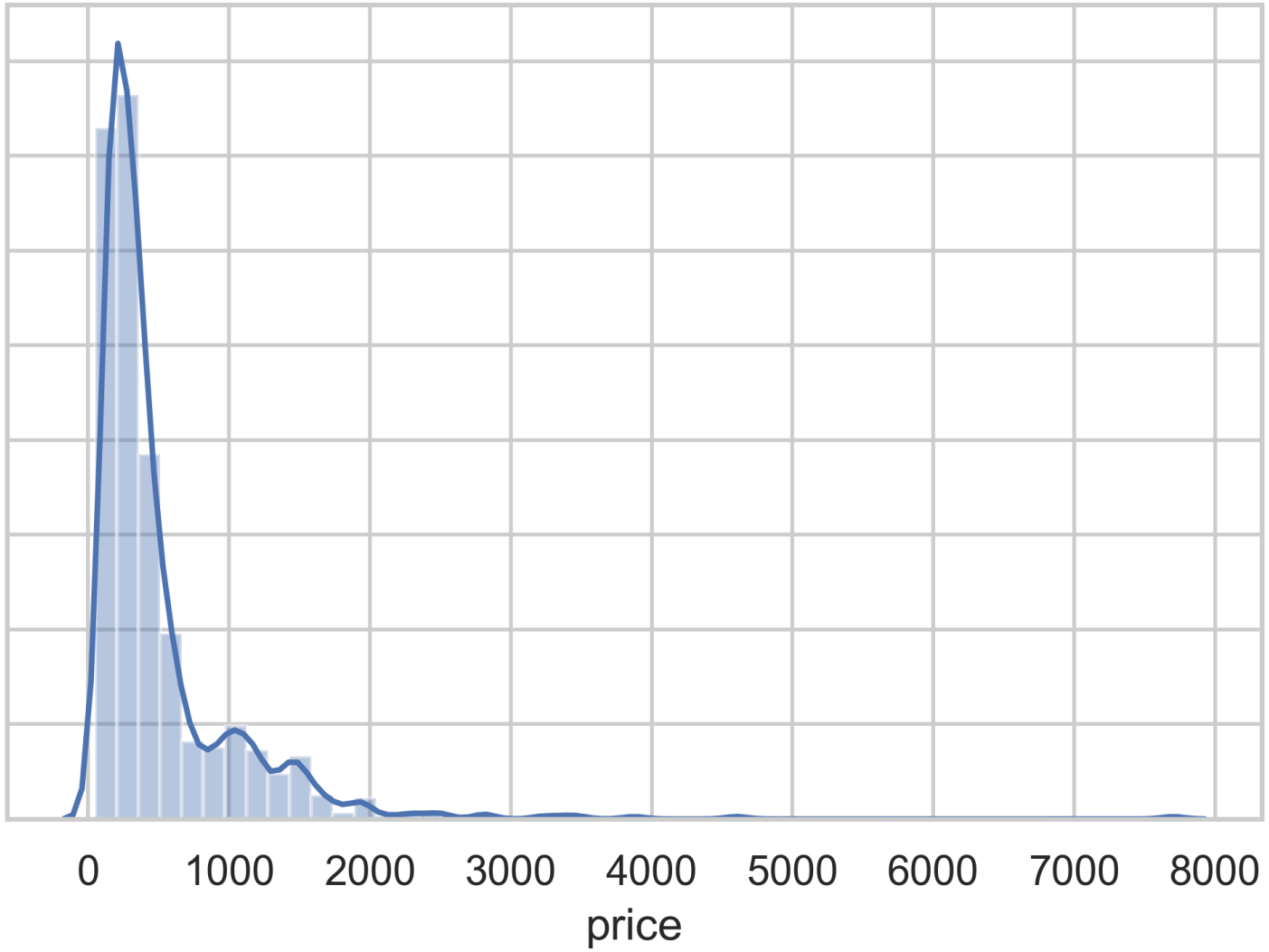
BASELINE MODEL

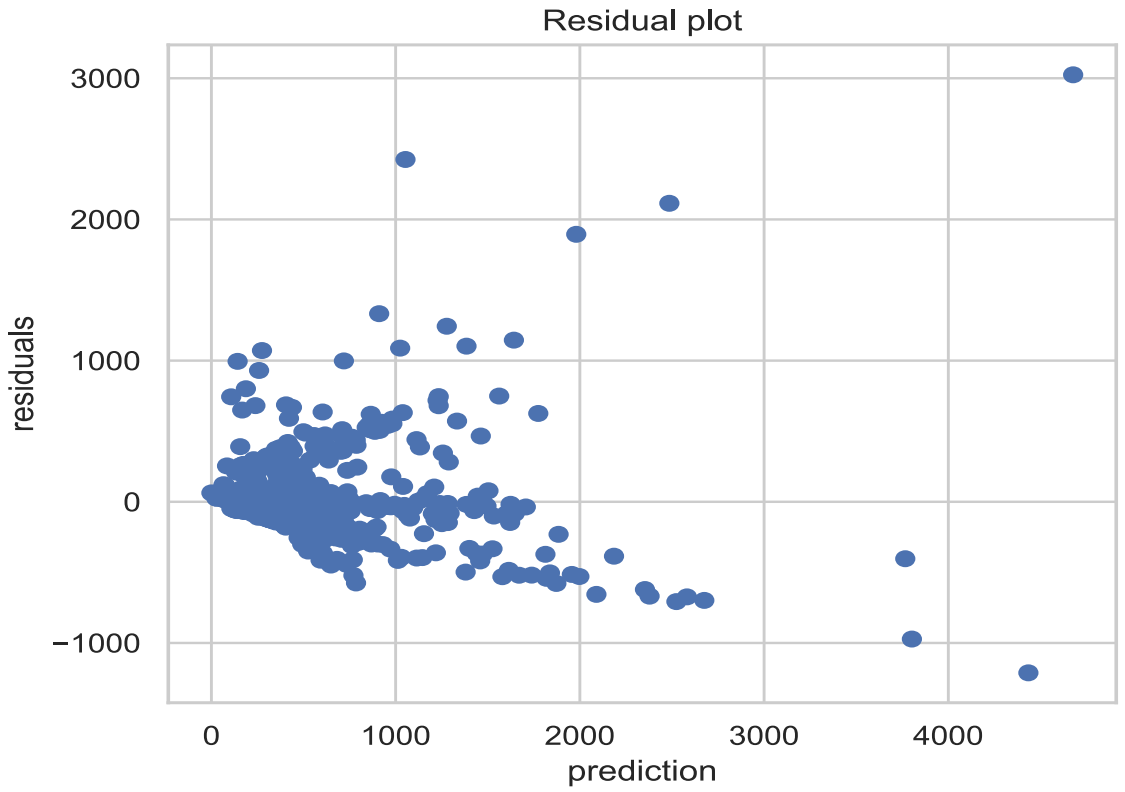


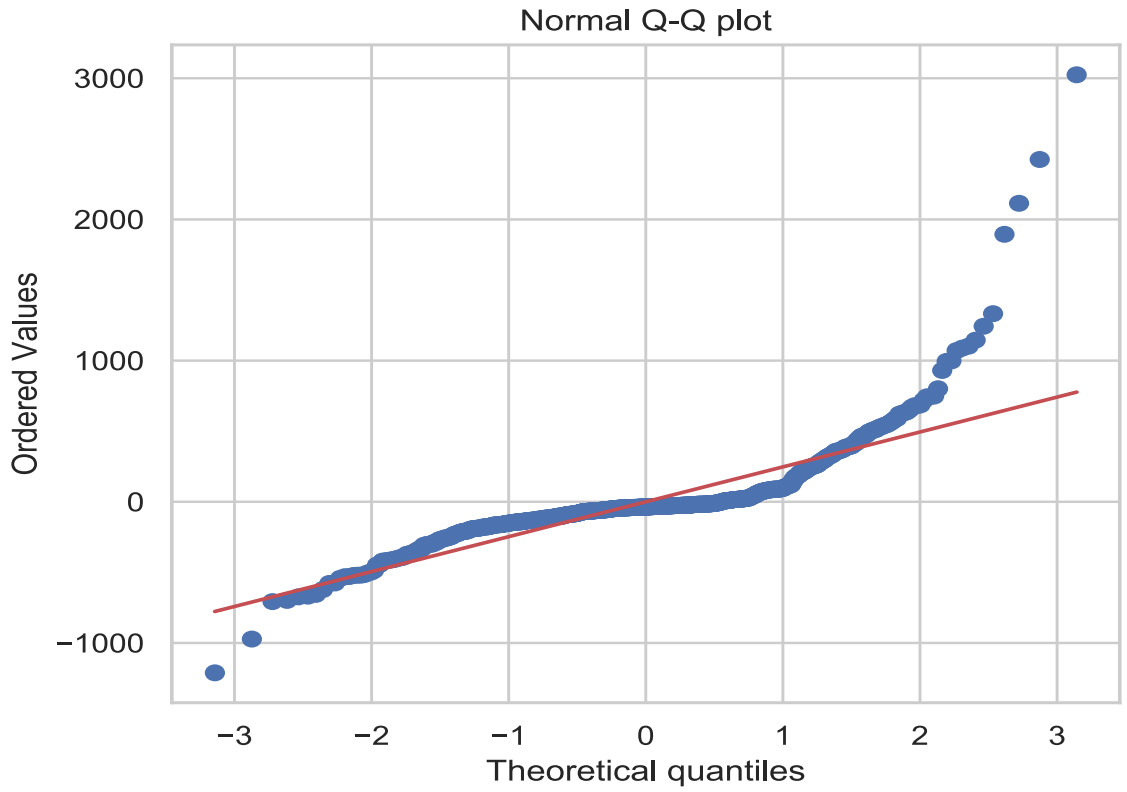
OBSERVATIONS

- Weight is correlated with the target.
- Net-weight and gross-weight are highly correlated with each other.
- Diamond pcs and carat are highly correlated with each other .

TARGET DISTRIBUTION




Residual plot
residuals
prediction
3000
2000
1000
0
−1000
0
1000
2000
3000
4000


Normal Q-Q plot
Ordered Values
Theoretical quantiles
3000
2000
1000
0
−1000
−3
−2
−1
0
1
2
3

Dep. Variable:	price	**R-squared:**	0.728
Model:	OLS	**Adj. R-squared:**	0.726
Method:	Least Squares	**F-statistic:**	439.9
Date:	Wed, 18 Sep 2019	**Prob (F-statistic):**	1.74e-229
Time:	21:06:18	**Log-Likelihood:**	-5881.2
No. Observations:	828	**AIC:**	1.177e+04
Df Residuals:	822	**BIC:**	1.180e+04
Df Model:	5		
Covariance Type:	nonrobust		

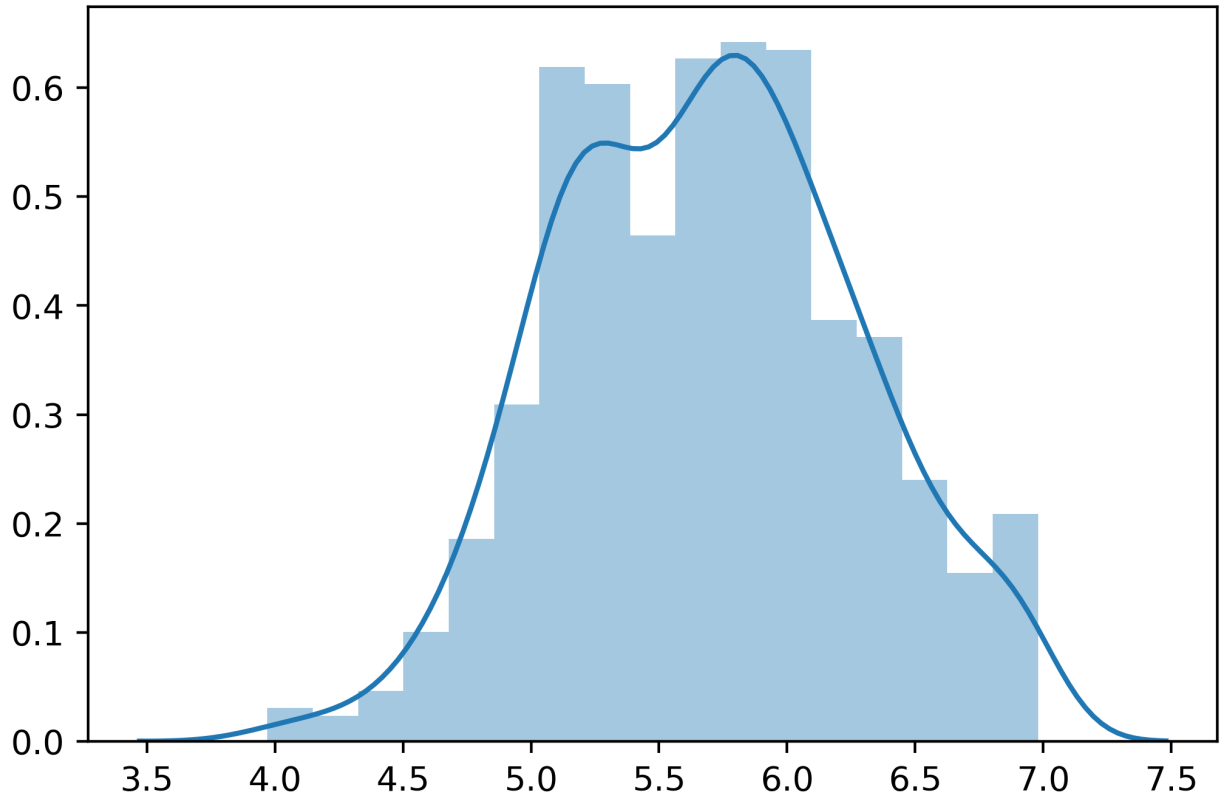
	coef	std err	t	P>|t|	[0.025	0.975]
const	-99.2502	31.437	-3.157	0.002	-160.957	-37.544
item_package_quantity	76.7709	23.482	3.269	0.001	30.679	122.863
gross_wieght	-12.4536	28.931	-0.430	0.667	-69.240	44.333
net_wieght	95.6652	29.796	3.211	0.001	37.179	154.151
carat_CT	1877.1648	113.884	16.483	0.000	1653.627	2100.702
diamond_pcs	0.5275	0.721	0.732	0.464	-0.887	1.942

Omnibus:	674.046	**Durbin-Watson:**	2.049
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	24121.791
Skew:	3.403	**Prob(JB):**	0.00
Kurtosis:	28.551	**Cond. No.**	301.

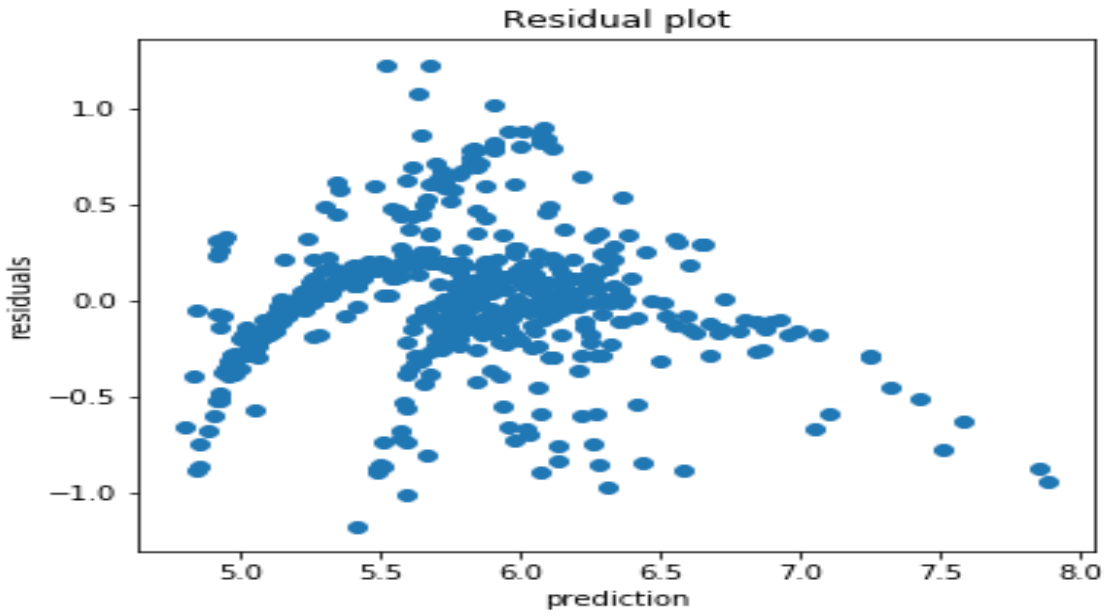
BASELINE MODEL SUMMARY

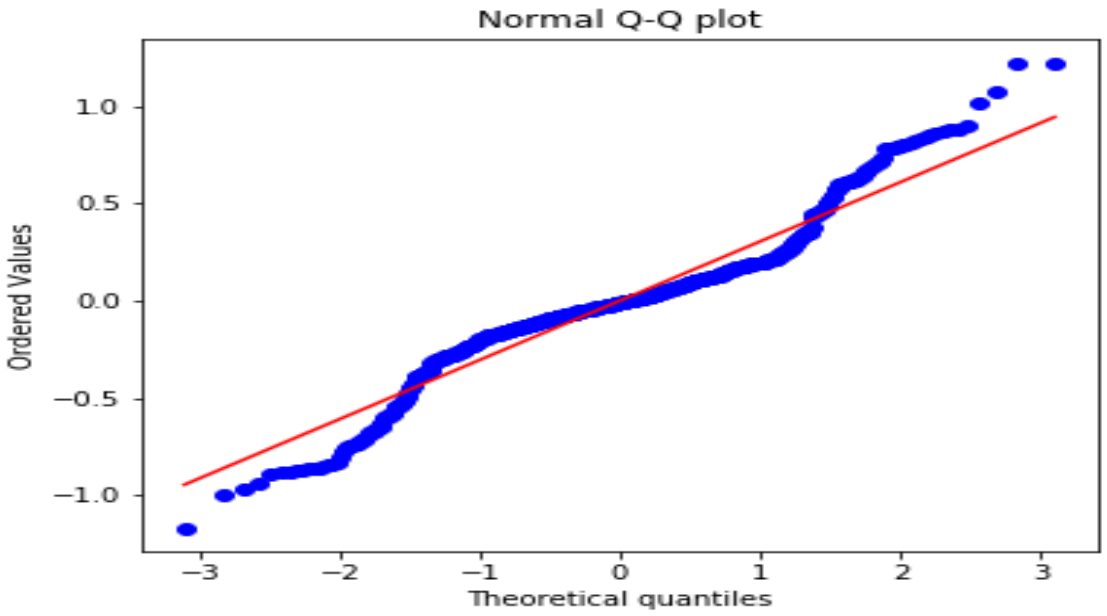
brand_Malabar	net_wieght	carat_CT	IGI_cert
0	2.940	0.00	0
1	2.580	0.00	0

SELECTED MODEL



TARGET DISTRIBUTION AFTER APPLYING LOG TRANSFORMATION ON IT


Residual plot
residuals
prediction
1.0
0.5
0.0
−0.5
−1.0
5.0
5.5
6.0
6.5
7.0
7.5
8.0


Normal Q-Q plot
Ordered Values
Theoretical quantiles
1.0
0.5
0.0
−0.5
−1.0
−3
−2
−1
0
1
2
3

OLS Regression Results

Dep. Variable:	price	**R-squared:**	0.712
Model:	OLS	**Adj. R-squared:**	0.711
Method:	Least Squares	**F-statistic:**	448.5
Date:	Thu, 19 Sep 2019	**Prob (F-statistic):**	2.21e-194
Time:	02:34:51	**Log-Likelihood:**	-192.31
No. Observations:	730	**AIC:**	394.6
Df Residuals:	725	**BIC:**	417.6
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	5.2811	0.032	165.915	0.000	5.219	5.344
brand_Malabar	-0.5269	0.029	-18.125	0.000	-0.584	-0.470
net_wieght	0.1544	0.005	31.321	0.000	0.145	0.164
carat_CT	2.0646	0.170	12.153	0.000	1.731	2.398
IGI_cert	0.1114	0.037	3.052	0.002	0.040	0.183

Omnibus:	45.812	**Durbin-Watson:**	1.892
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	168.648
Skew:	0.115	**Prob(JB):**	2.39e-37
Kurtosis:	5.344	**Cond. No.**	64.7

MODEL SUMMARY ON TRANSFORMED DATA

TESTING SCORE : 0.68

CONCLUSIONS AND DISCUSSION