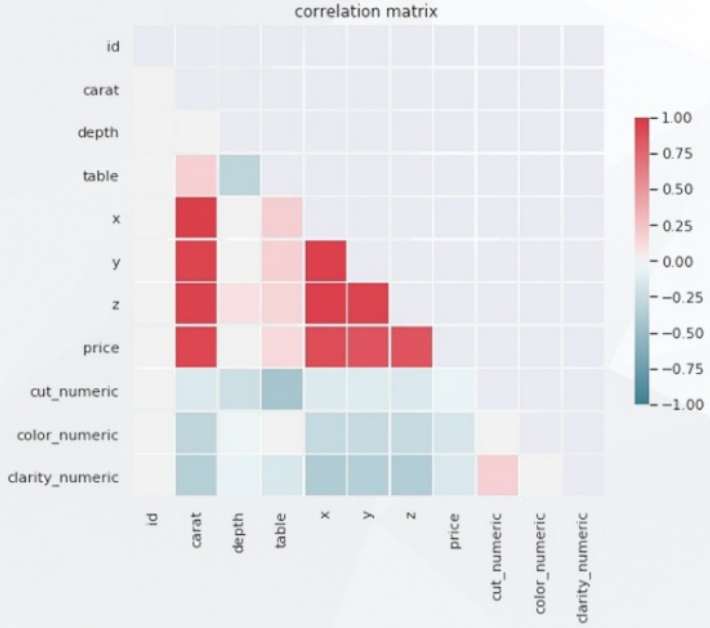
kaggle

LIMPIEZA Y STANDARDIZAZION DE LOS DATOS

	carat	cut_numeric	color_numeric	clarity_numeric	depth	table	x	y	z
0	0.30	3	5	4	62.8	56.0	4.29	4.31	2.70
1	0.34	5	6	3	62.6	55.0	4.46	4.49	2.80
2	0.40	3	7	3	60.3	62.0	4.70	4.75	2.85
3	0.40	4	3	5	61.8	59.2	4.72	4.74	2.92
4	0.90	3	7	3	61.0	63.0	6.10	6.13	3.73

	carat	cut_numeric	color_numeric	clarity_numeric	depth	table	x	y	z
0	-1.051355	-0.810564	0.352161	-0.033632	0.736126	-0.650176	-1.283716	-1.236438	-1.180531
1	-0.966764	0.982588	0.940769	-0.641579	0.596363	-1.097476	-1.132102	-1.080061	-1.039596
2	-0.839879	-0.810564	1.529377	-0.641579	-1.010919	2.033618	-0.918059	-0.854183	-0.969129
3	-0.839879	0.086012	-0.825055	0.574315	0.037308	0.781181	-0.900222	-0.862870	-0.870475
4	0.217504	-0.810564	1.529377	-0.641579	-0.521746	2.480917	0.330526	0.344710	0.271096

MATRIX DE CORRELACION



METRICA DE LOS MODELOS

1- GradientBoostingRegressor

tuned whit `params=(n_estimators = 100, loss='huber', min_samples_leaf=6, max_depth=7,max_leaf_nodes=250, min_samples_split=45)`

`RMSE=474.89712806494185`

2- HistGradientBoostingRegressor

tuned whit `params=(max_depth=8, min_samples_leaf=6)`

`RMSE=531.4433490228398`

3- RandomForestRegressor

tuned whit `params=(n_estimators=500, max_depth=15, min_samples_leaf=3, n_jobs=-1)`

`RMSE=533.875343545384`

4- Extra Tree Regressor

tuned whit `params=(max_features='log2', n_jobs=-1)`

`RMSE=546.6463913185166`

5- BaggingRegressor

tuned whit `params=(n_estimators=14)`

`RMSE=547.4550303837045`

6- KNeighborsClassifier

tuned whit `params=(n_neighbors=1, algorithm='brute', p=1)`

`RMSE=773.0013101770397`

7- Linear Regression

tuned whit `params=(without params)`

`RMSE=1165.9673289710706`

8- GaussianNB

tuned whit `params=(without params)`

`RMSE=1341.8201798533694`

GRAFICOS DE LOS DOS PRIMEROS MODELOS