Real Estate Pricing Model

	feature	original_feature	engineering	type	r2_score
0	neighborhood_mode_imputed	neighborhood	mode_imputed	categorical	0.575692
1	overall_qual_mean_imputed^2	overall_qual	mean_imputed^2	numeric	0.54345
2	exter_qual_mode_imputed	exter_qual	mode_imputed	categorical	0.522003
3	bsmt_qual_na_val_imputed	bsmt_qual	na_val_imputed	categorical	0.510579
4	kitchen_qual_mode_imputed	kitchen_qual	mode_imputed	categorical	0.504355
5	gr_liv_area_mean_imputed	gr_liv_area	mean_imputed	numeric	0.483091
6	garage_cars_mean_imputed^2	garage_cars	mean_imputed^2	numeric	0.403627
7	1st_flr_sf_mean_imputed	1st_flr_sf	mean_imputed	numeric	0.371196

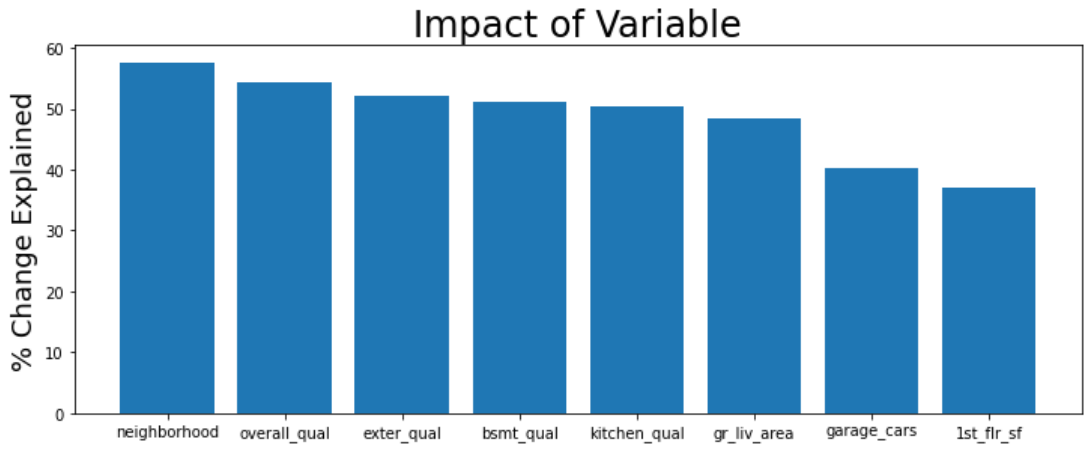
Main Model Summary

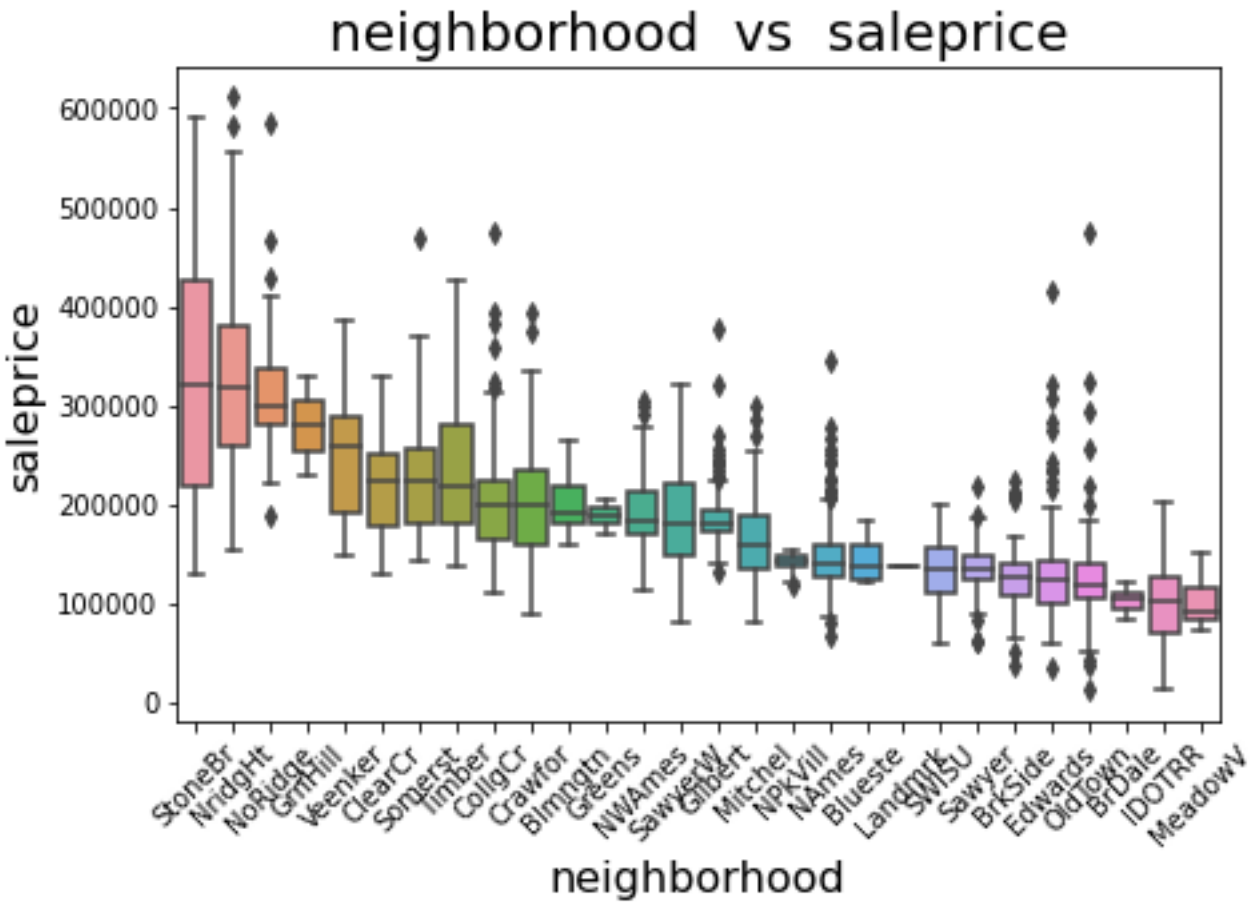
Dep. Variable:	saleprice	R-squared (uncentered):	0.969
Model:	OLS	Adj. R-squared (uncentered):	0.969
Method:	Least Squares	F-statistic:	8032.
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	0.00
Time:	00:39:14	Log-Likelihood:	-24356.
No. Observations:	2051	AIC:	4.873e+04
Df Residuals:	2043	BIC:	4.877e+04
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
neighborhood_mode_imputed	0.1845	0.020	9.103	0.000	0.145	0.224
overall_qual_mean_imputed^2	1664.8980	76.191	21.852	0.000	1515.478	1814.318
exter_qual_mode_imputed	-0.0017	0.024	-0.072	0.943	-0.049	0.045
bsmt_qual_na_val_imputed	0.0128	0.021	0.606	0.545	-0.029	0.054
kitchen_qual_mode_imputed	0.0743	0.022	3.335	0.001	0.031	0.118
gr_liv_area_mean_imputed	24.0781	1.959	12.292	0.000	20.237	27.920
garage_cars_mean_imputed^2	5078.6186	364.248	13.943	0.000	4364.283	5792.954
1st_flr_sf_mean_imputed	13.9753	2.418	5.780	0.000	9.234	18.717

Omnibus:	654.421	Durbin-Watson:	1.990
Prob(Omnibus):	0.000	Jarque-Bera (JB):	11255.446
Skew:	1.042	Prob(JB):	0.00
Kurtosis:	14.286	Cond. No.	1.72e+05

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[3] The condition number is large, 1.72e+05. This might indicate that there are
strong multicollinearity or other numerical problems.





Variable Type: categorical

Engineering: mode_imputed

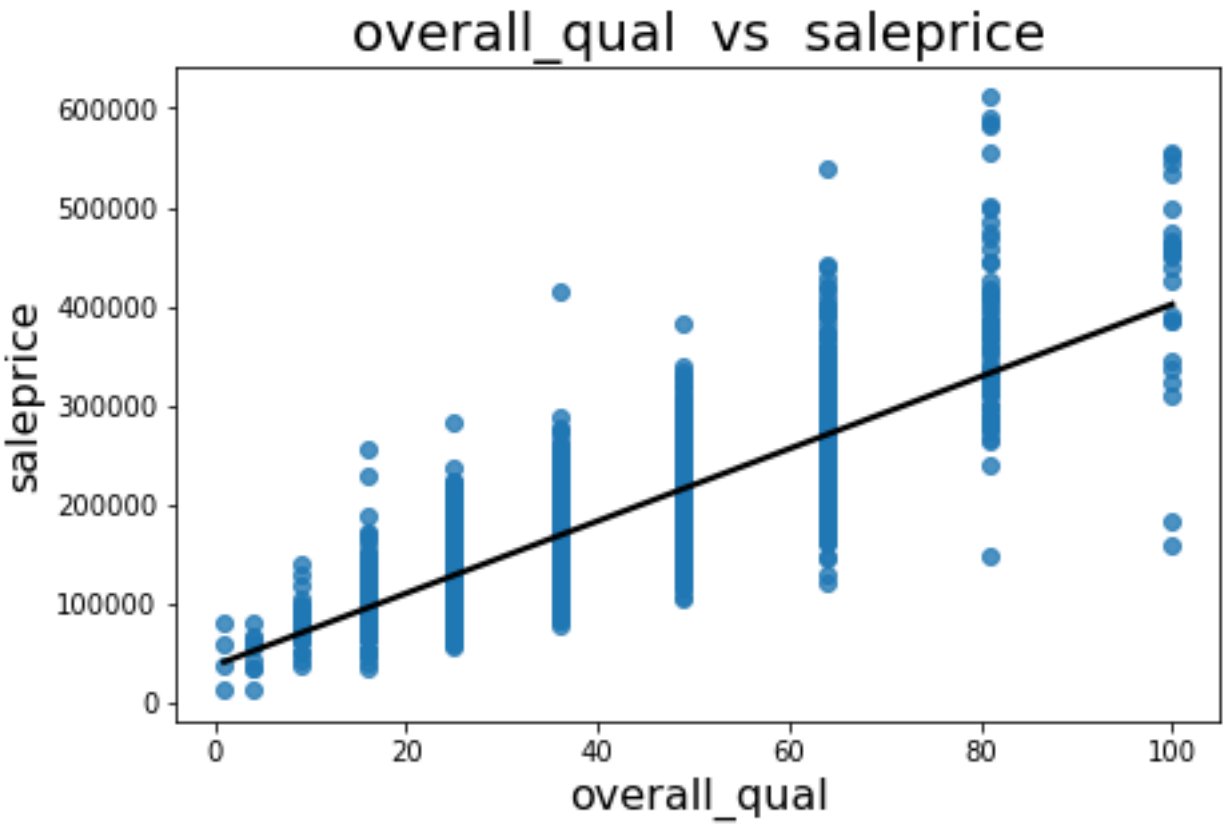
OLS Regression Results

Dep. Variable:	neighborhood_mode_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:15	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.9071	0.005	167.685	0.000	0.896	0.918

Omnibus:	464.669	Durbin-Watson:	1.840
Prob(Omnibus):	0.000	Jarque-Bera (JB):	2479.441
Skew:	-0.962	Prob(JB):	0.00
Kurtosis:	8.031	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: numeric

Engineering: mean_imputed^2

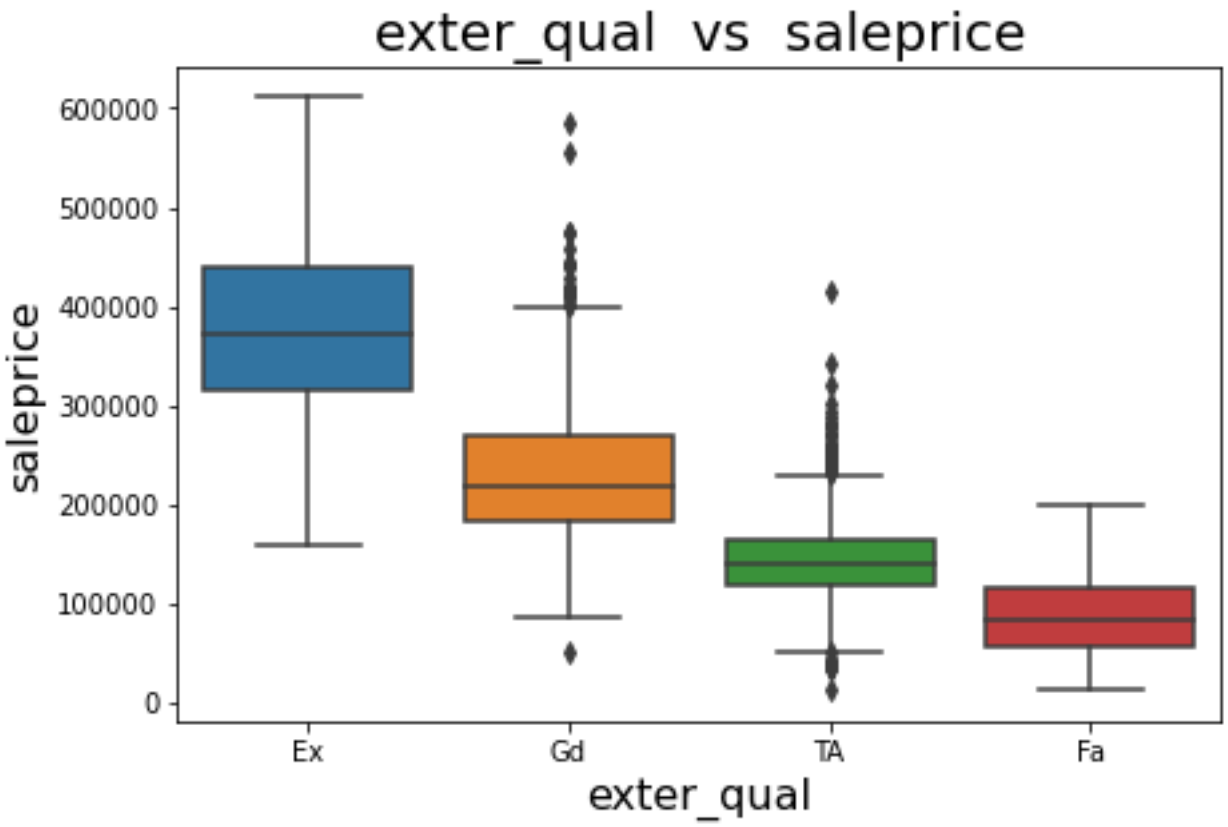
OLS Regression Results

Dep. Variable:	overall_qual_mean_imputed^2	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:16	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.0002	1.16e-06	183.527	0.000	0.000	0.000

Omnibus:	144.248	Durbin-Watson:	2.017
Prob(Omnibus):	0.000	Jarque-Bera (JB):	685.946
Skew:	0.094	Prob(JB):	1.12e-149
Kurtosis:	5.827	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: categorical

Engineering: mode_imputed

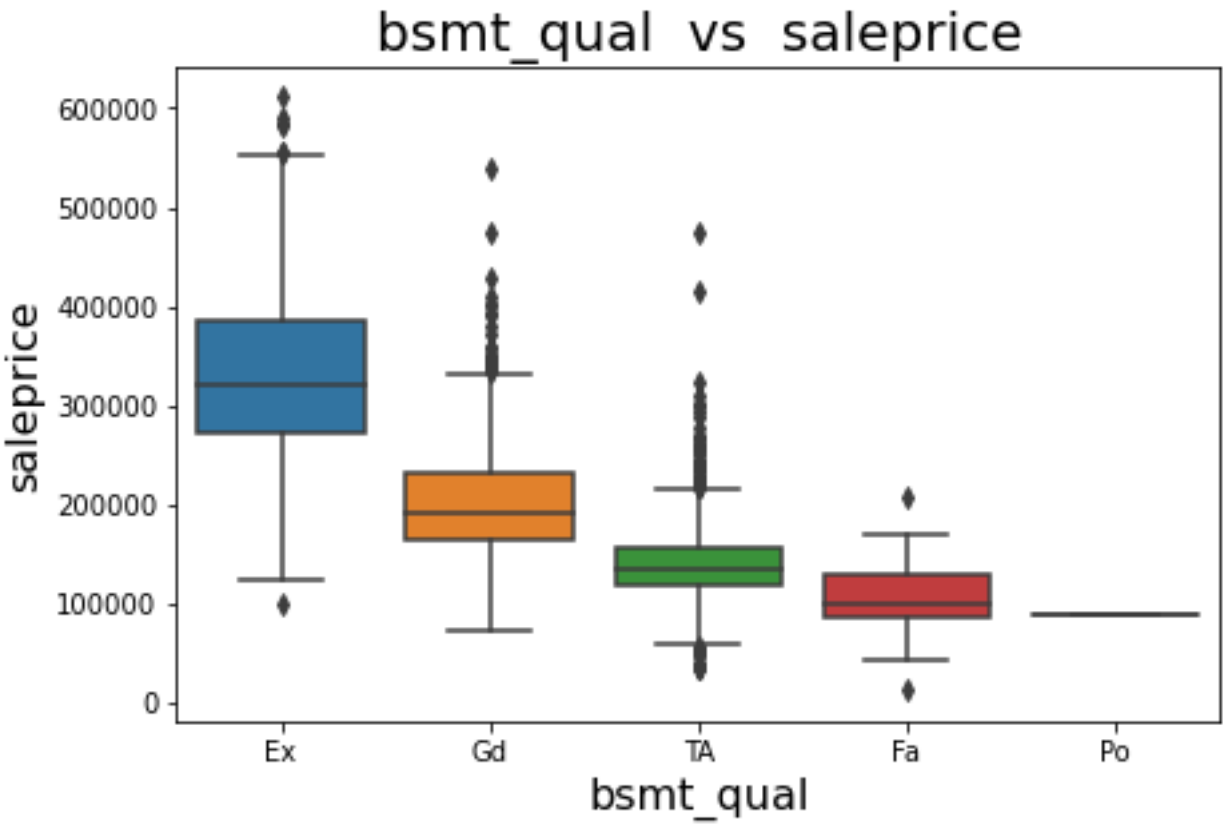
OLS Regression Results

Dep. Variable:	exter_qual_mode_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:16	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.8871	0.006	157.264	0.000	0.876	0.898

Omnibus:	349.967	Durbin-Watson:	1.851
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1262.876
Skew:	-0.815	Prob(JB):	5.89e-275
Kurtosis:	6.481	Cond. No.	1.00

Notes:
[1] R² is computed without centering (uncentered) since the model does not contair
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: categorical

Engineering: na_val_imputed

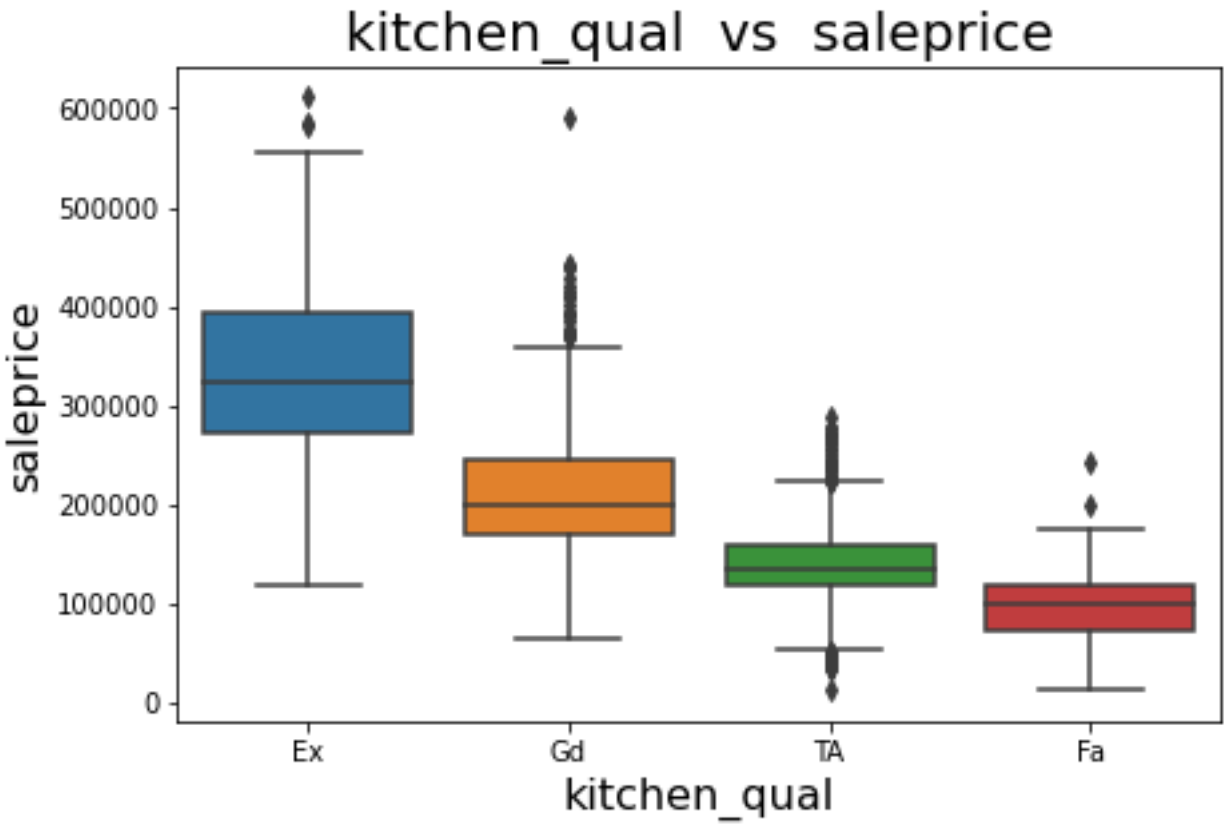
OLS Regression Results

Dep. Variable:	bsmt_qual_na_val_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:17	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.8893	0.006	151.150	0.000	0.878	0.901

Omnibus:	319.233	Durbin-Watson:	1.796
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1266.431
Skew:	-0.714	Prob(JB):	9.95e-276
Kurtosis:	6.575	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: categorical

Engineering: mode_imputed

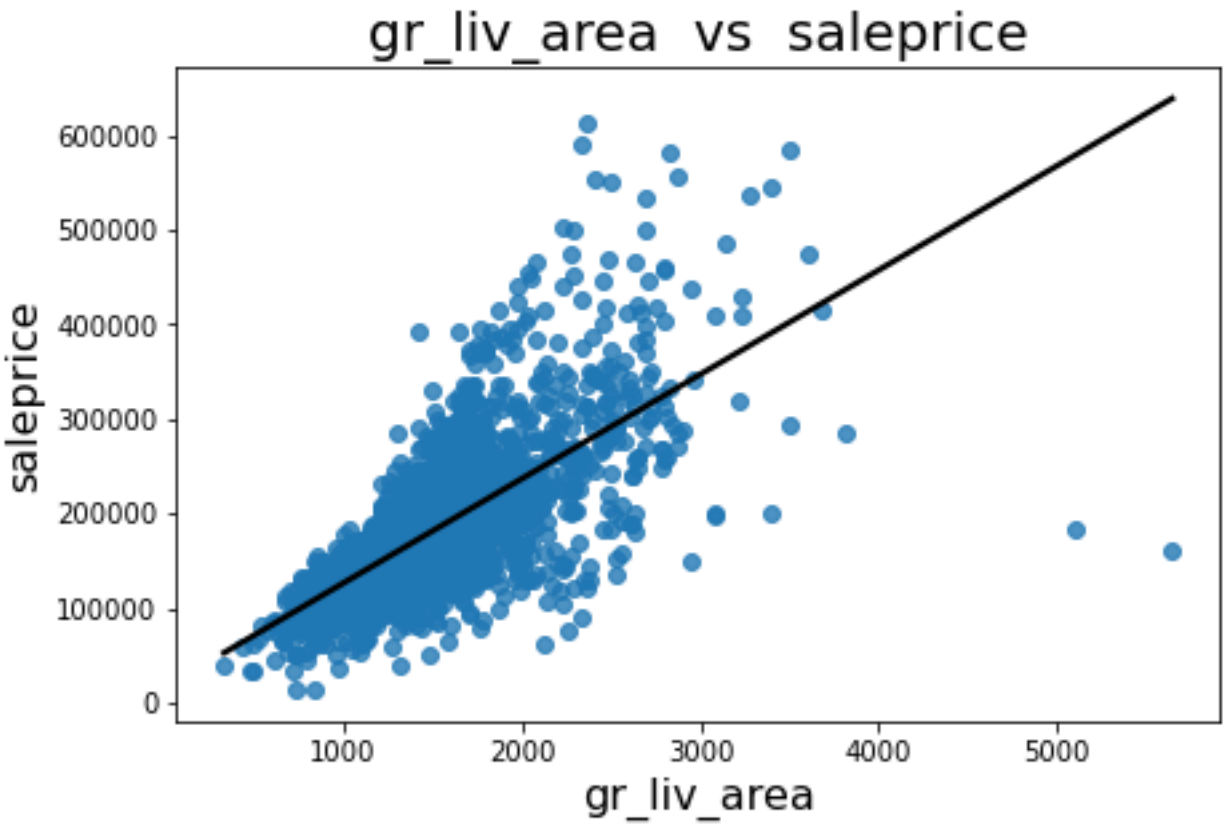
OLS Regression Results

Dep. Variable:	kitchen_qual_mode_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:17	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.8853	0.006	154.161	0.000	0.874	0.897

Omnibus:	286.940	Durbin-Watson:	1.856
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1054.602
Skew:	-0.659	Prob(JB):	9.91e-230
Kurtosis:	6.256	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: numeric

Engineering: mean_imputed

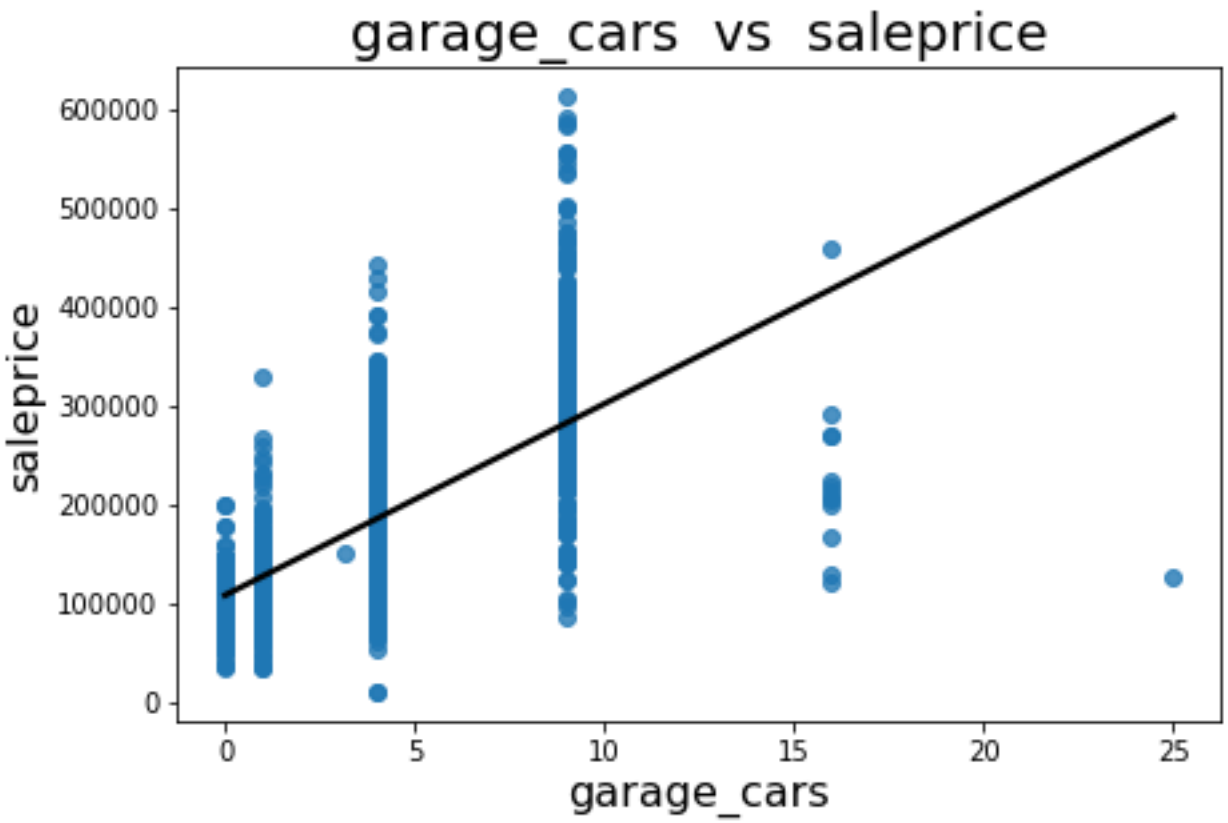
OLS Regression Results

Dep. Variable:	gr_liv_area_mean_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:18	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.0076	5.08e-05	150.514	0.000	0.008	0.008

Omnibus:	371.964	Durbin-Watson:	1.917
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6622.773
Skew:	0.303	Prob(JB):	0.00
Kurtosis:	11.782	Cond. No.	1.00

Notes:
[1] R^2 is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: numeric

Engineering: mean_imputed^2

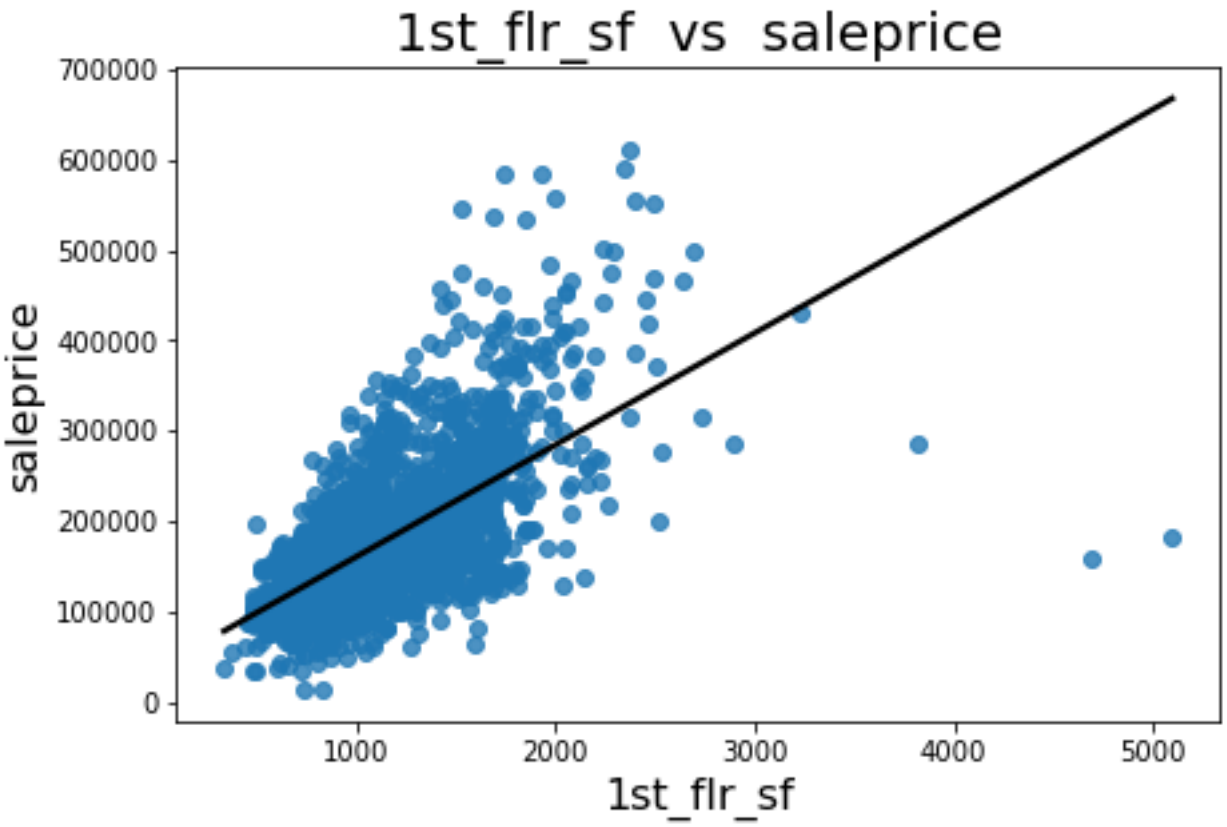
OLS Regression Results

Dep. Variable:	garage_cars_mean_imputed^2	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:18	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	2.098e-05	2.28e-07	91.924	0.000	2.05e-05	2.14e-05

Omnibus:	1077.312	Durbin-Watson:	1.916
Prob(Omnibus):	0.000	Jarque-Bera (JB):	15473.080
Skew:	2.133	Prob(JB):	0.00
Kurtosis:	15.762	Cond. No.	1.00

Notes:
[1] R² is computed without centering (uncentered) since the model does not contair
[2] Standard Errors assume that the covariance matrix of the errors is correctly s



Variable Type: numeric

Engineering: mean_imputed

OLS Regression Results

Dep. Variable:	1st_flr_sf_mean_imputed	R-squared (uncentered):	
Model:	OLS	Adj. R-squared (uncentered):	
Method:	Least Squares	F-statistic:	
Date:	Fri, 14 Oct 2022	Prob (F-statistic):	
Time:	00:39:19	Log-Likelihood:	
No. Observations:	2051	AIC:	
Df Residuals:	2050	BIC:	
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
saleprice	0.0059	4.39e-05	133.906	0.000	0.006	0.006

Omnibus:	471.598	Durbin-Watson:	1.935
Prob(Omnibus):	0.000	Jarque-Bera (JB):	9453.482
Skew:	0.557	Prob(JB):	0.00
Kurtosis:	13.458	Cond. No.	1.00

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain
[2] Standard Errors assume that the covariance matrix of the errors is correctly s