House Price Prediction

Ankit Nigam

GCDAI – August Batch 2019

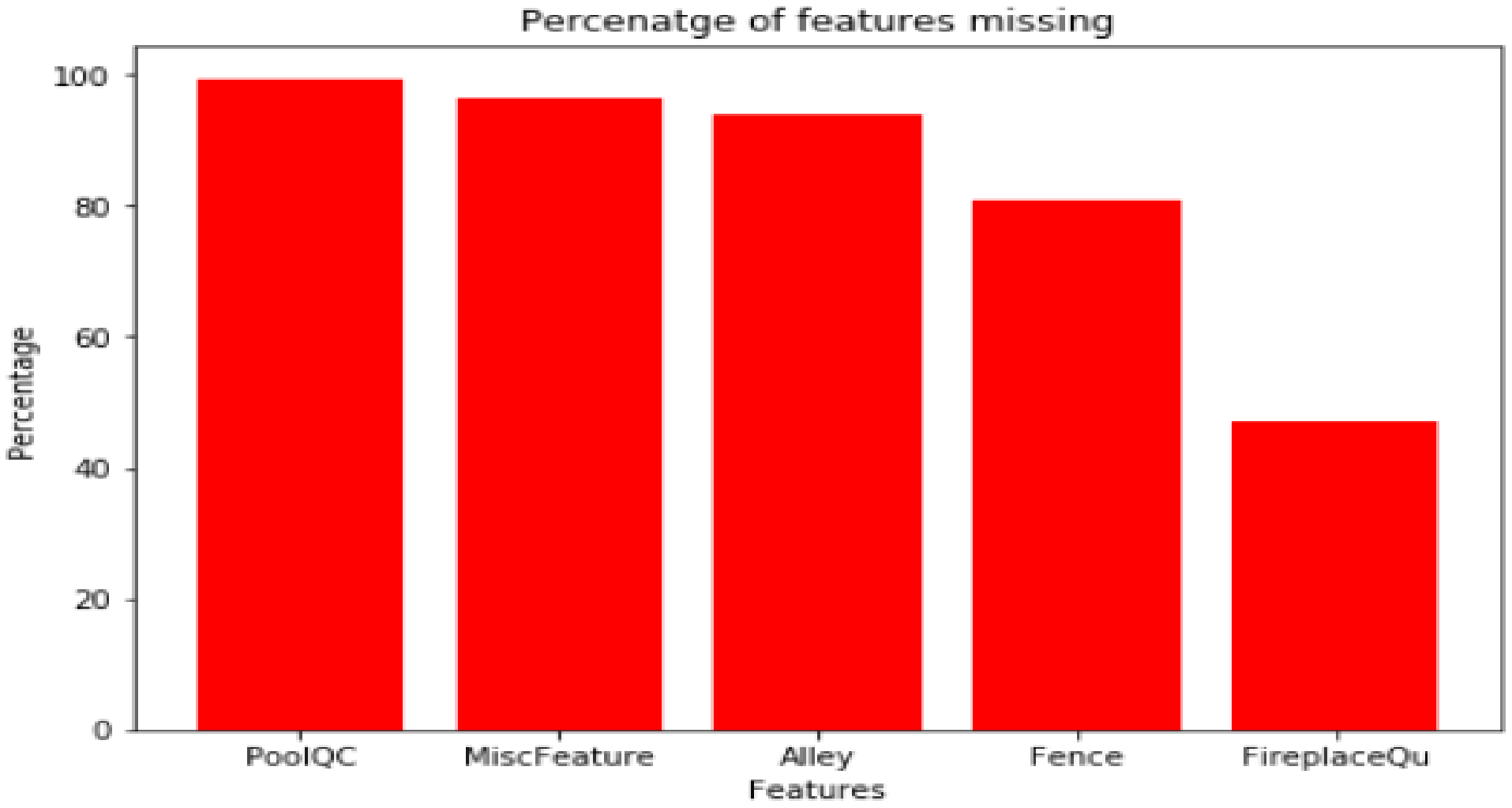
Dataset Information

https://github.com/insaid2018/Term-2/blob/master/Projects/houseprices.txt

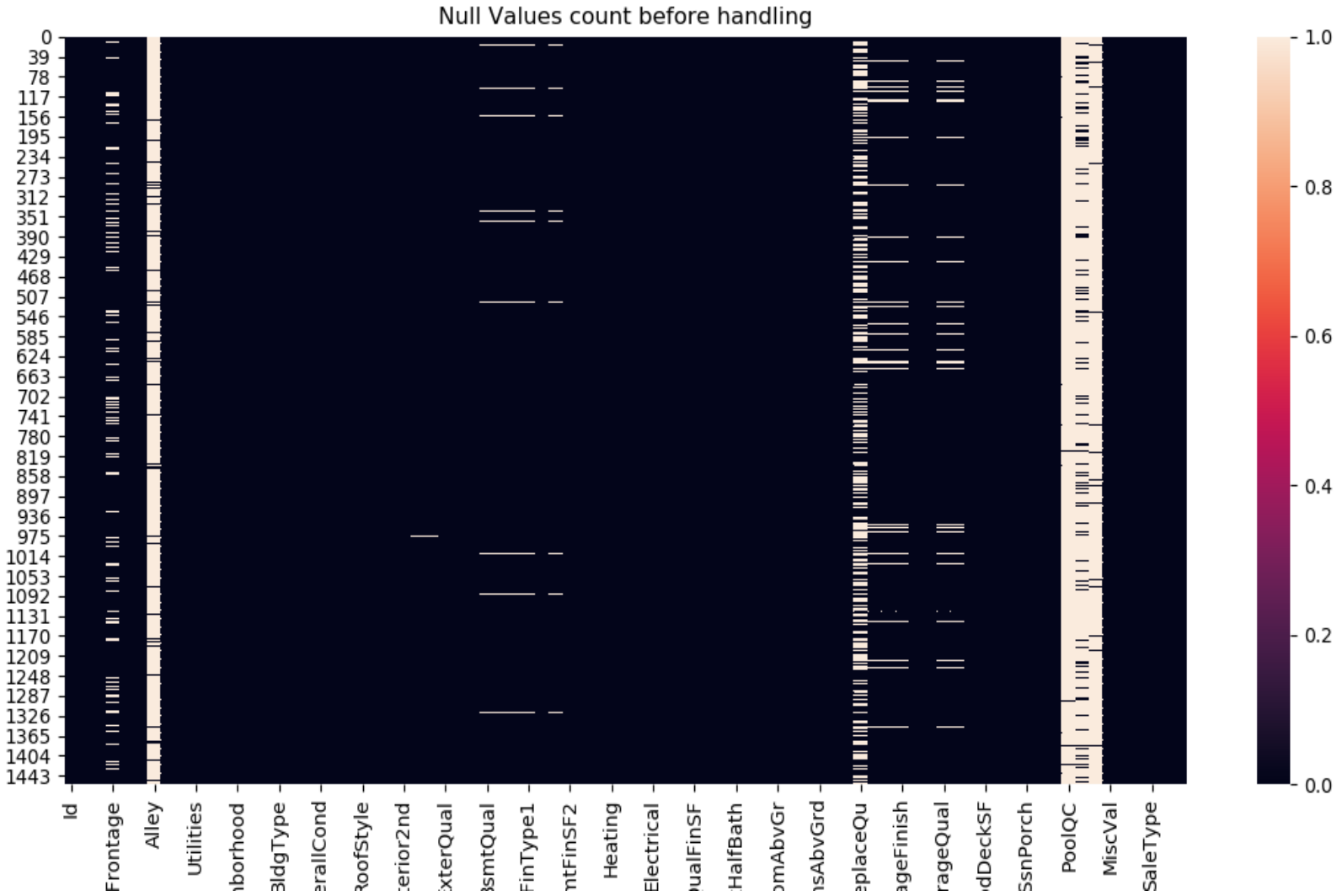
Target variable	Meaning
SalePrice	the property's sale price in dollars.

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood
1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	Inside	Gtl	CollgCr
2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	FR2	Gtl	Veenker

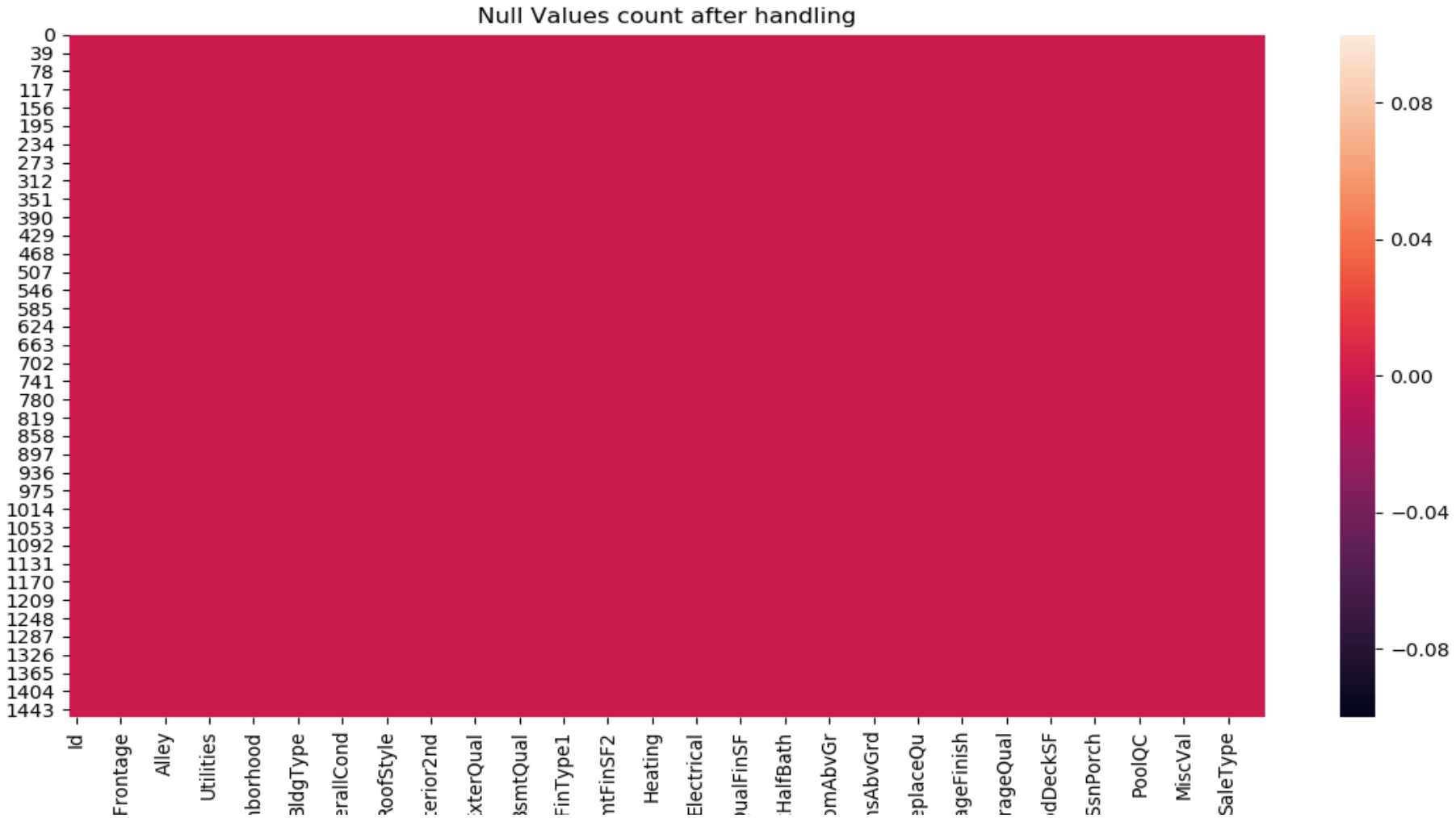
Handling missing values in dataset



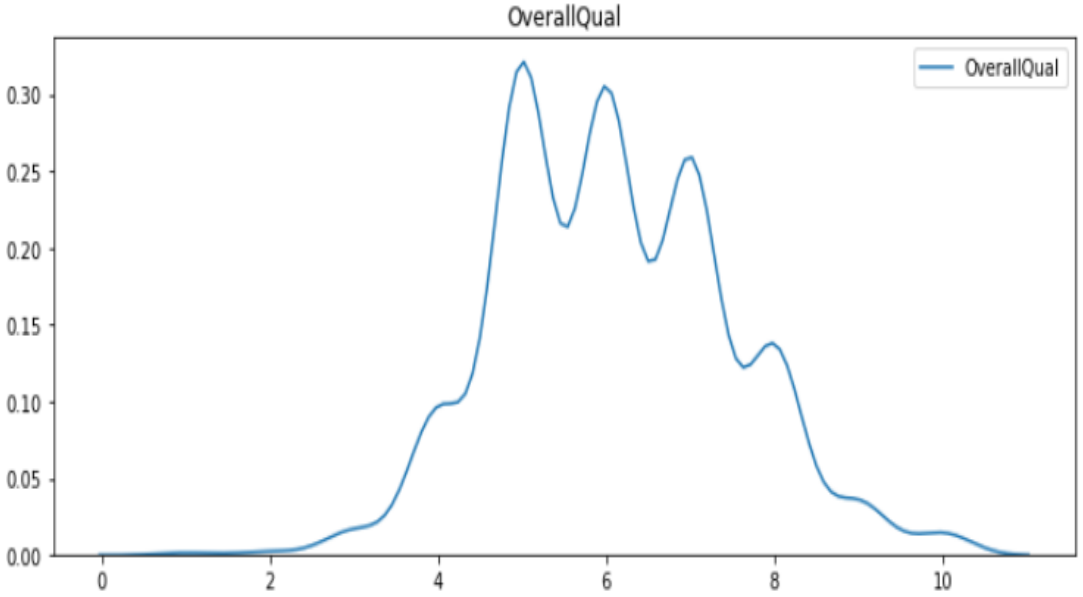
Handling missing values in dataset

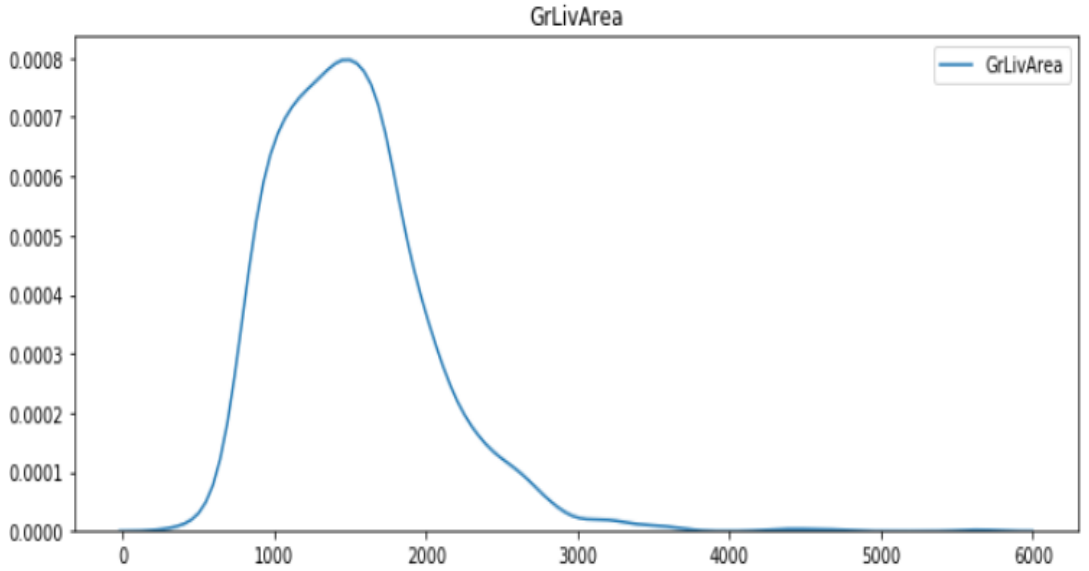


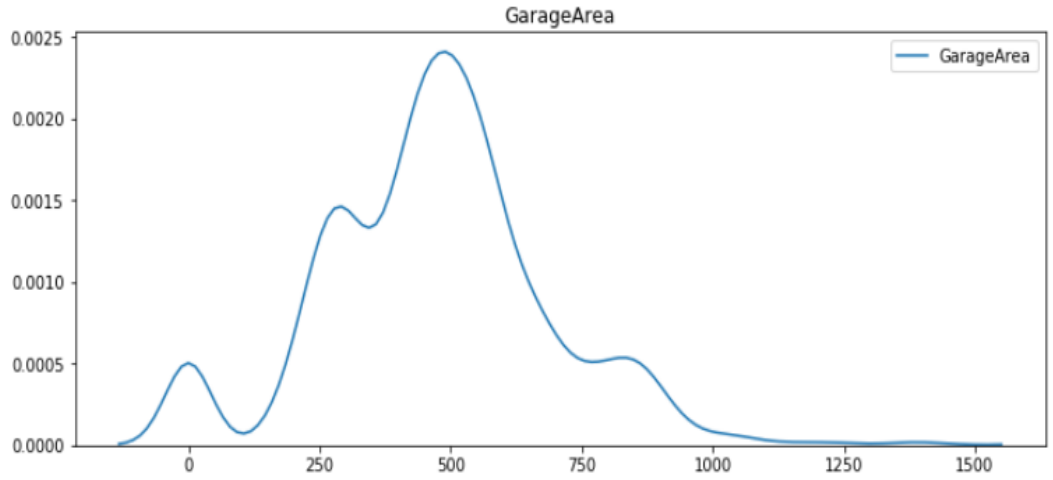
Handling missing values in dataset

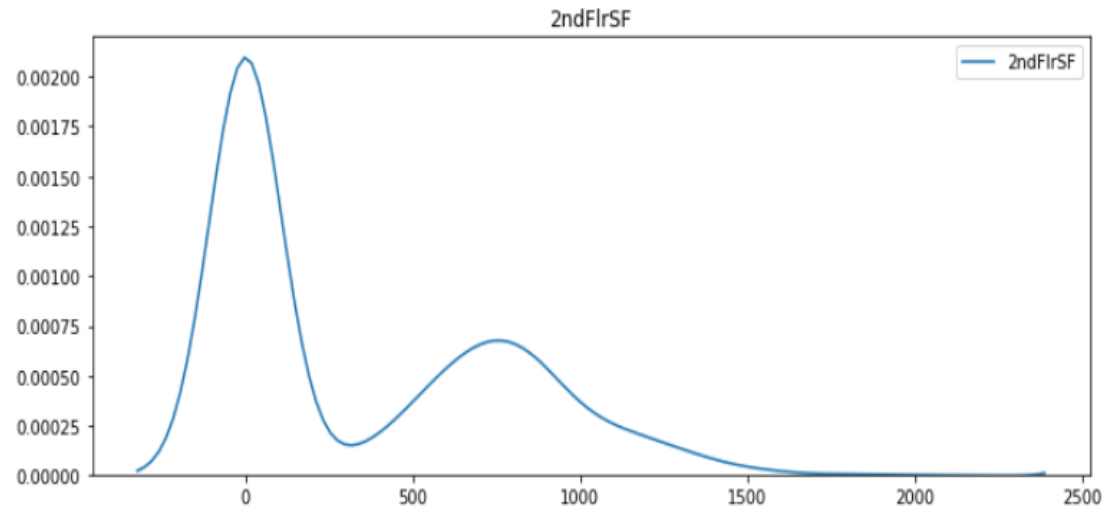


Distribution of input features

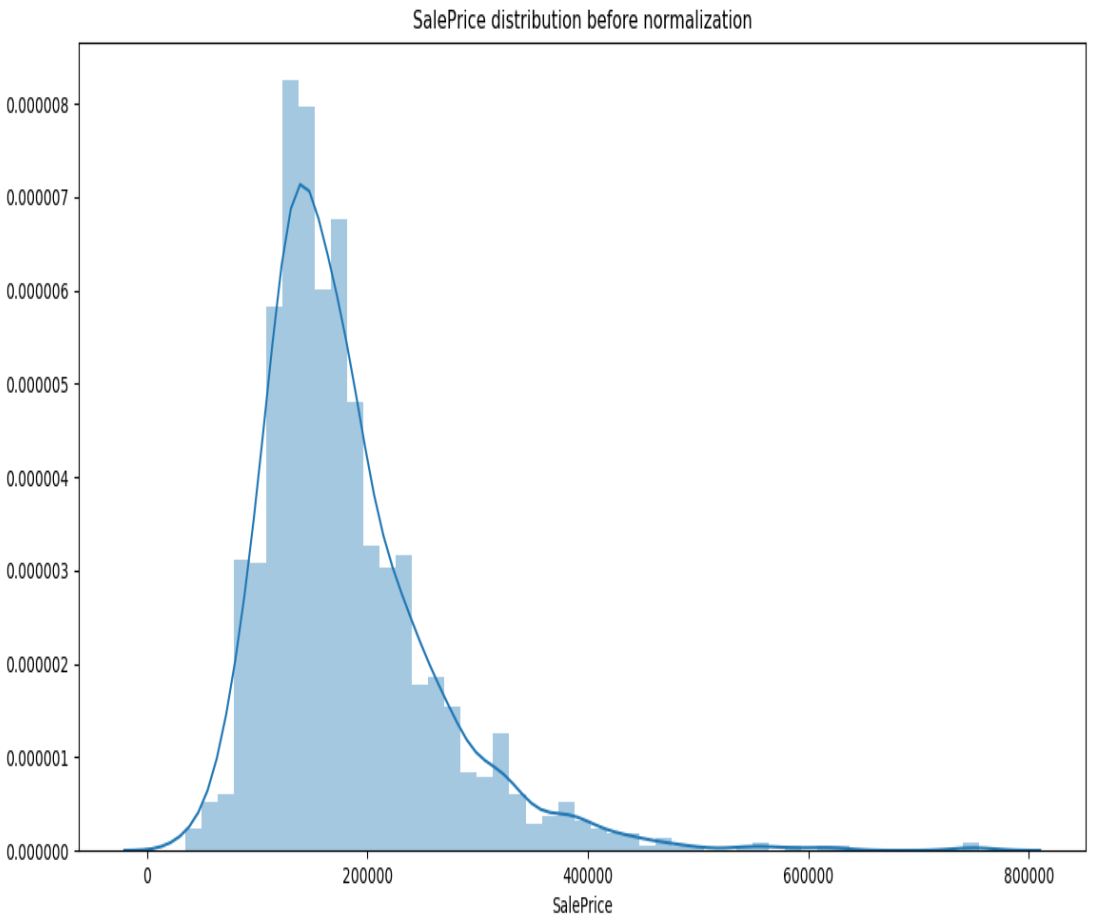


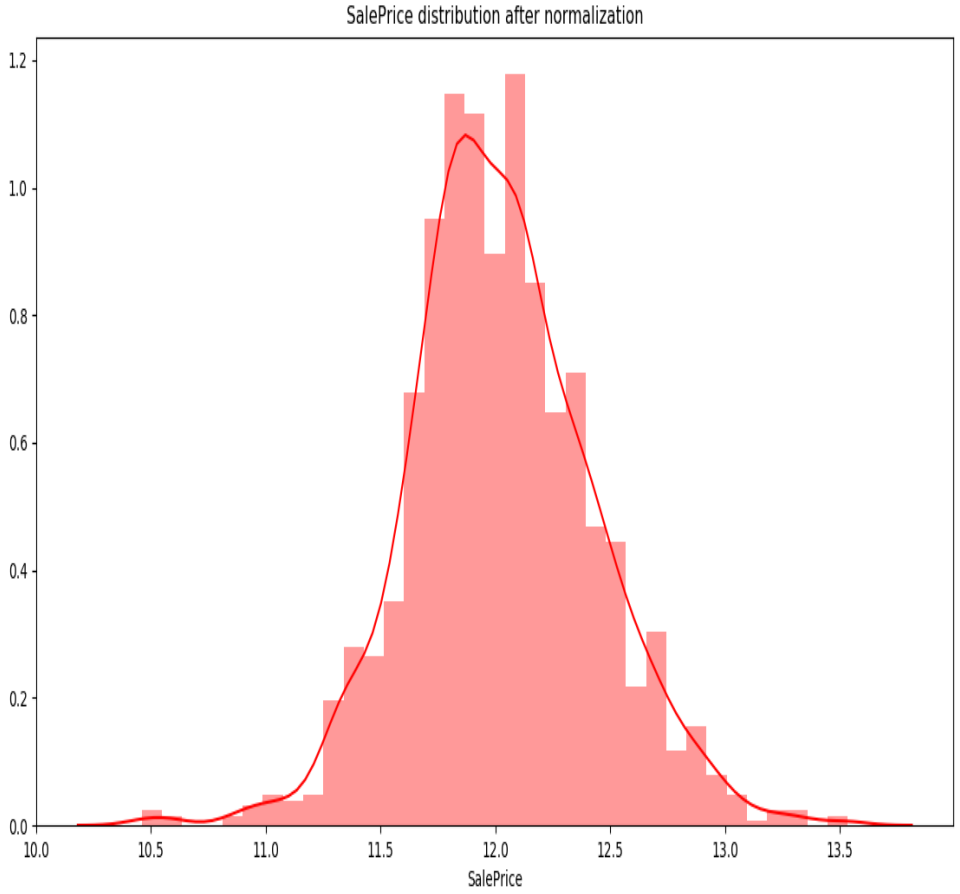




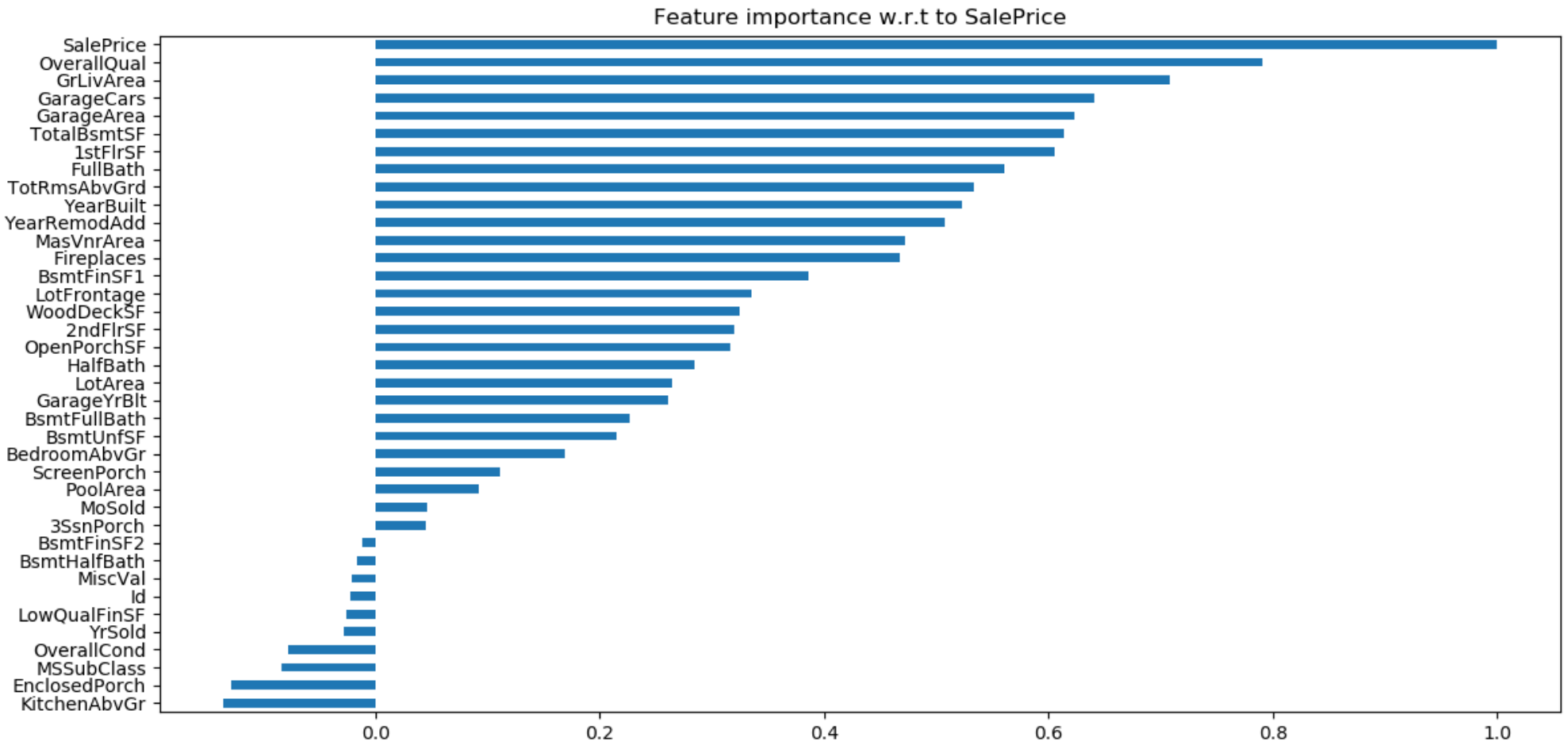


Distribution of target data

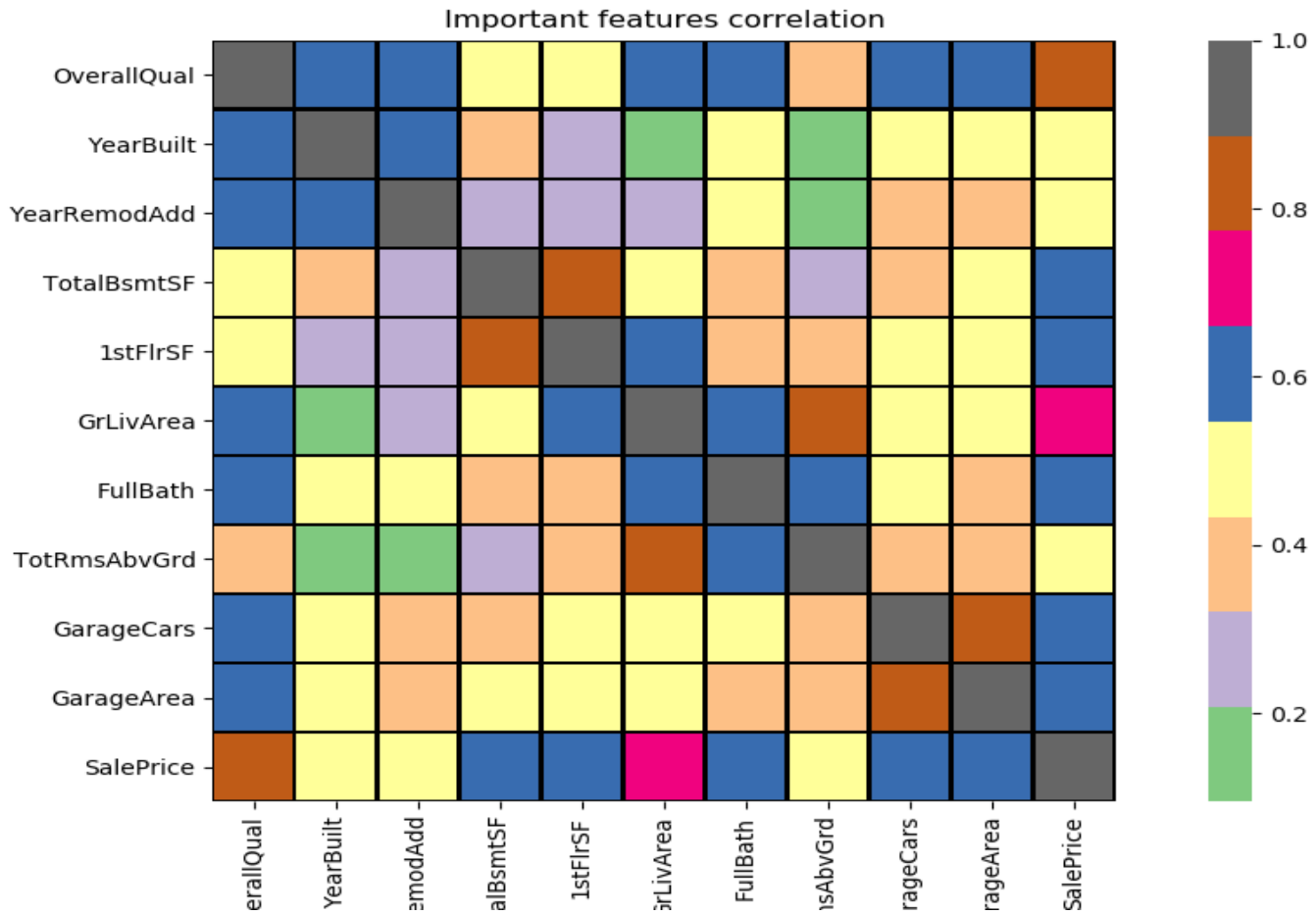




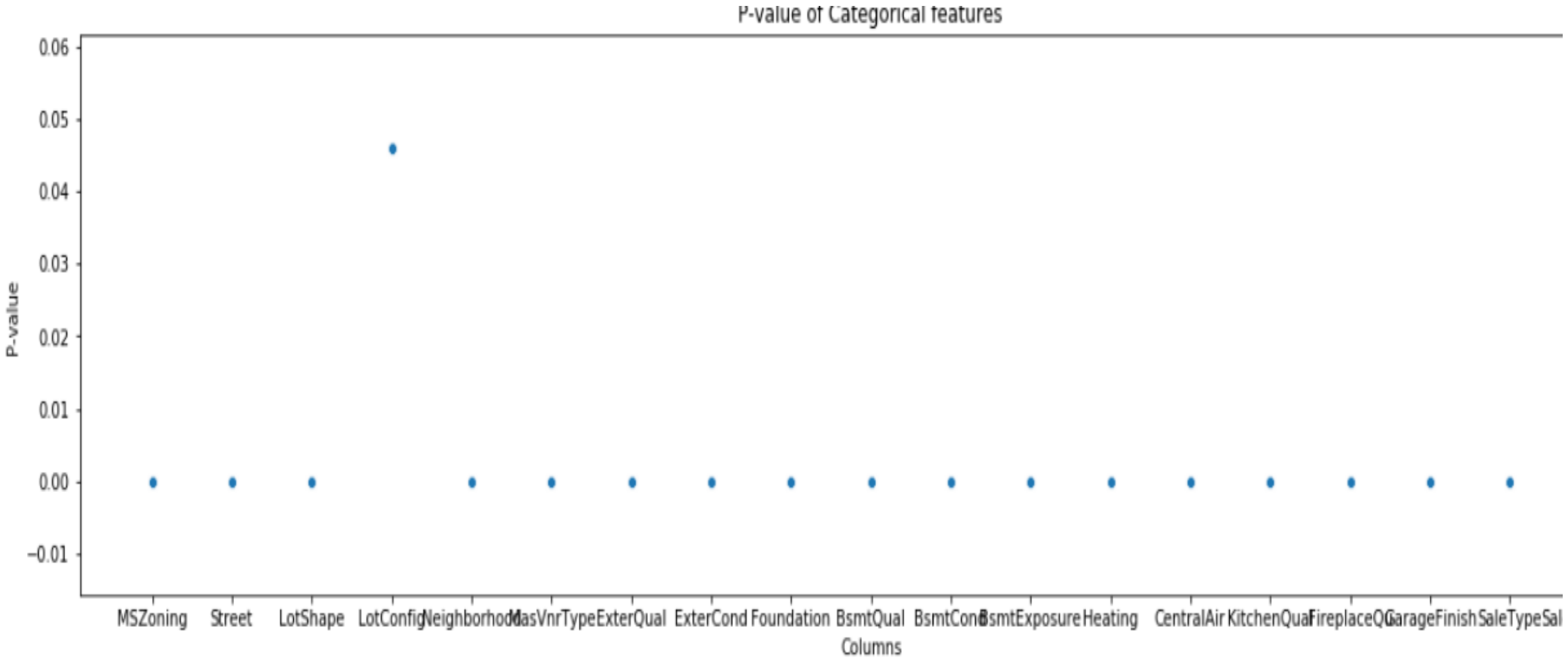
Selecting continuous features for prediction



Selecting continuous features for prediction

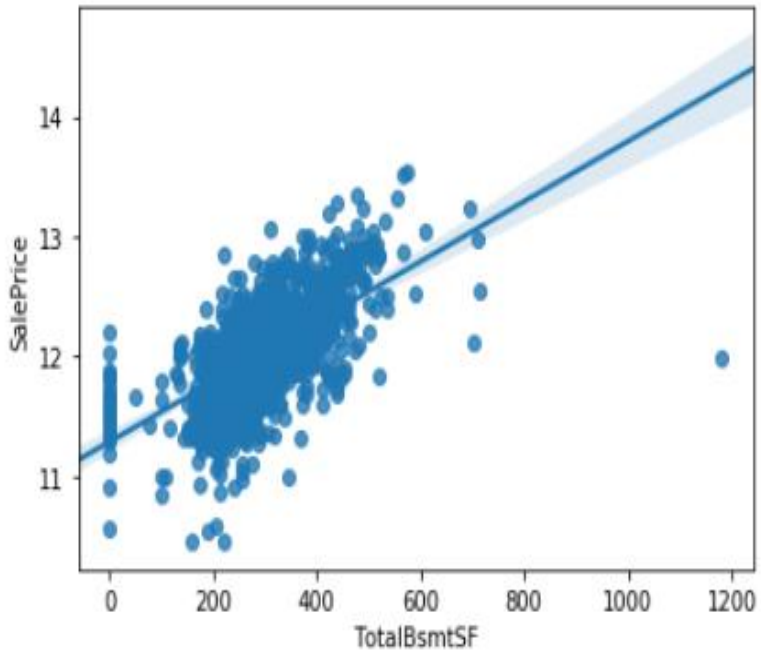


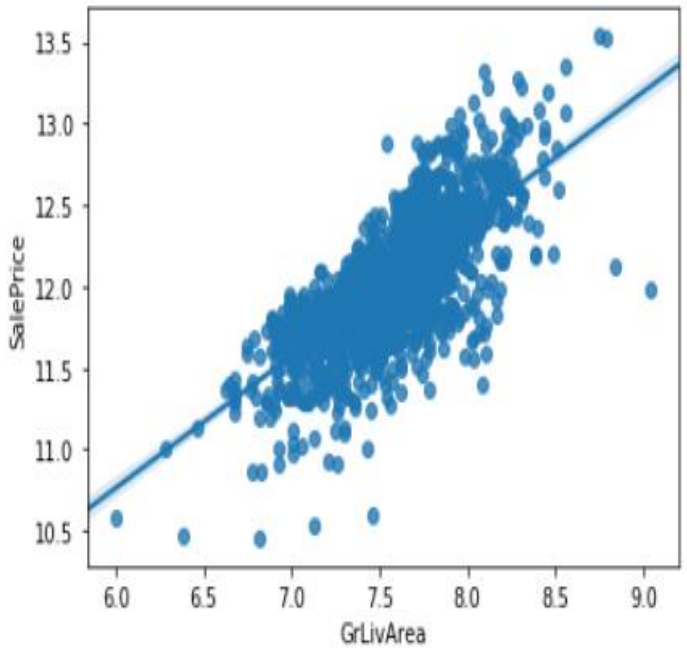
Selecting categorical features for prediction

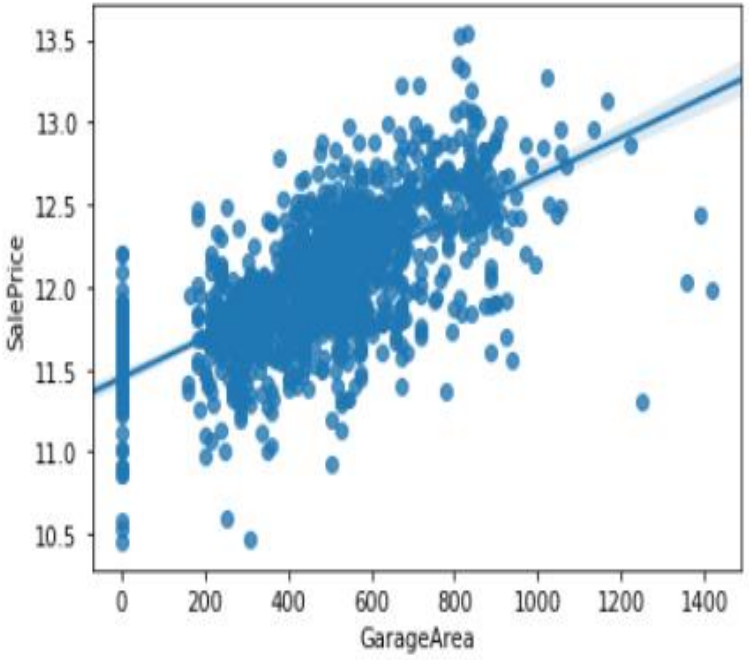


Assumptions for Linear Regression

- **There has to be linear relationship between input features and target variable**

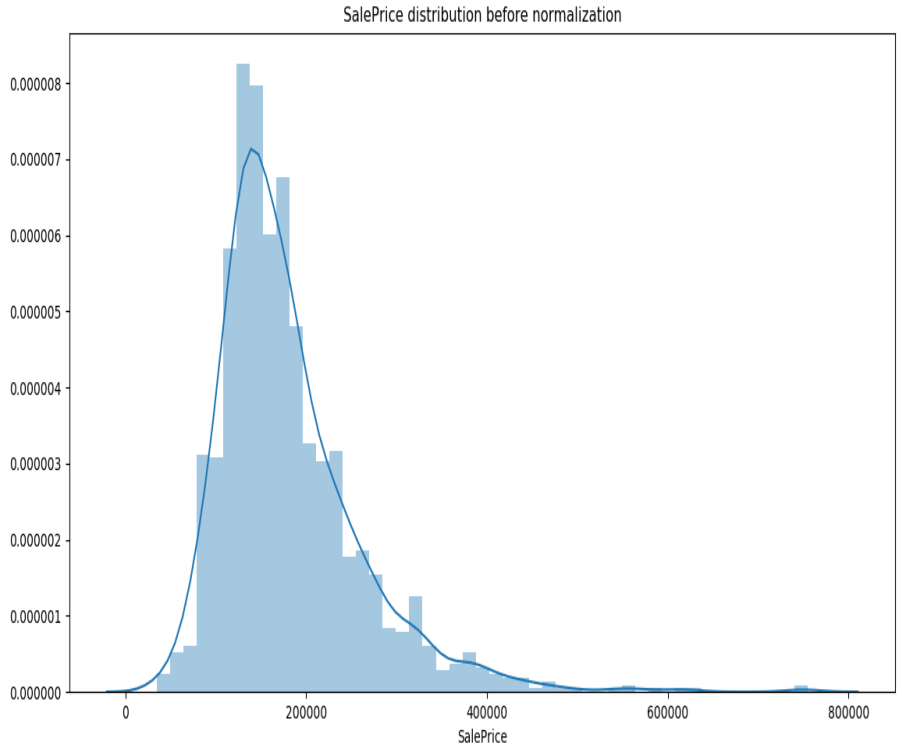


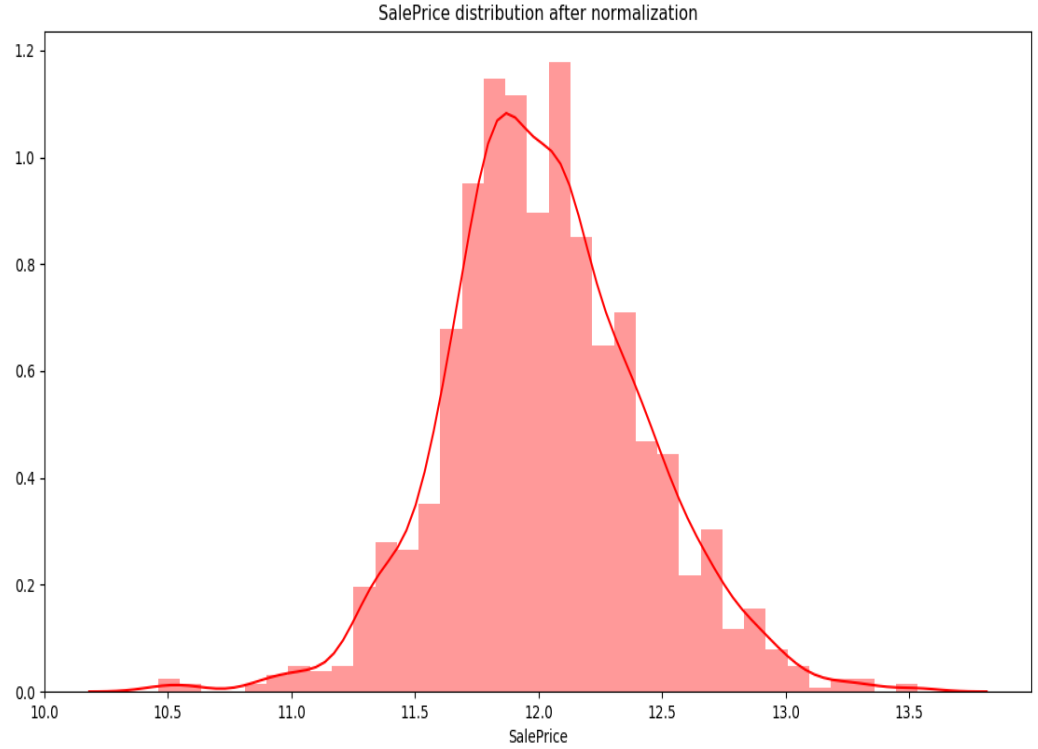




Assumptions for Linear Regression

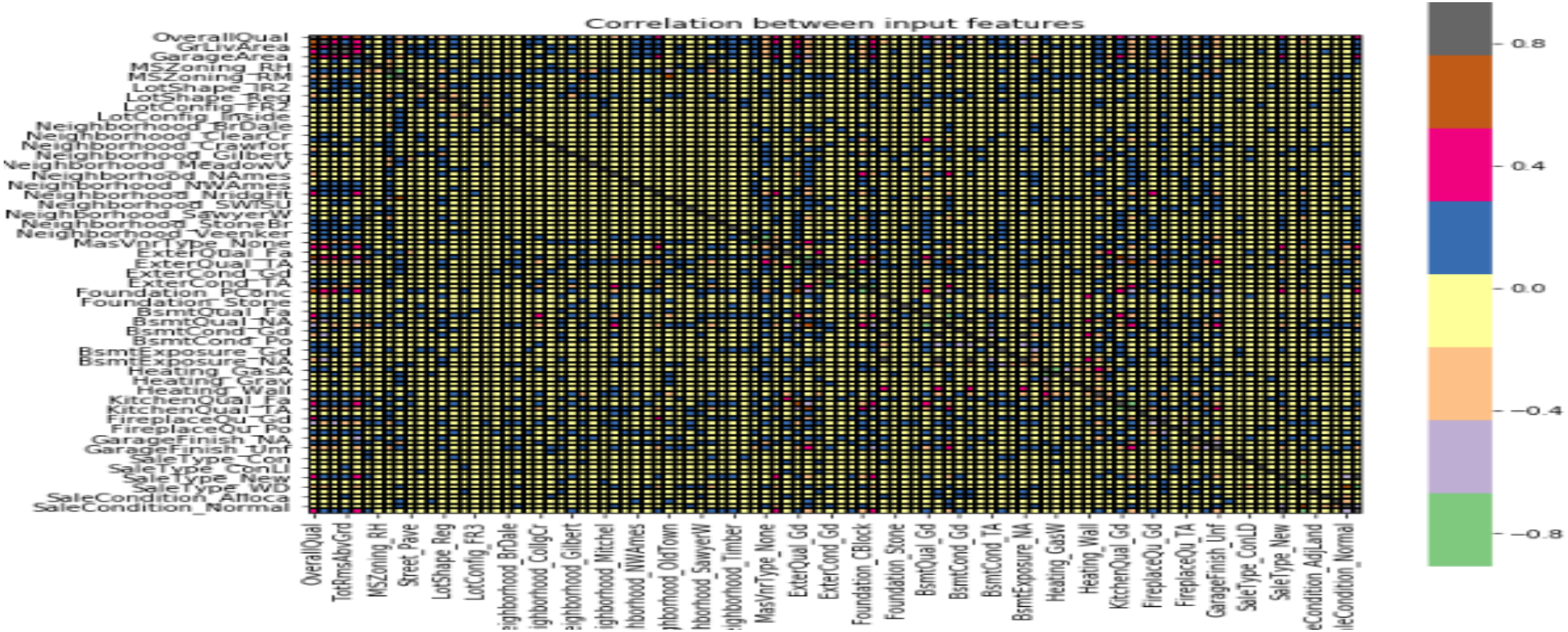
- **Target variable should be normally distributed**





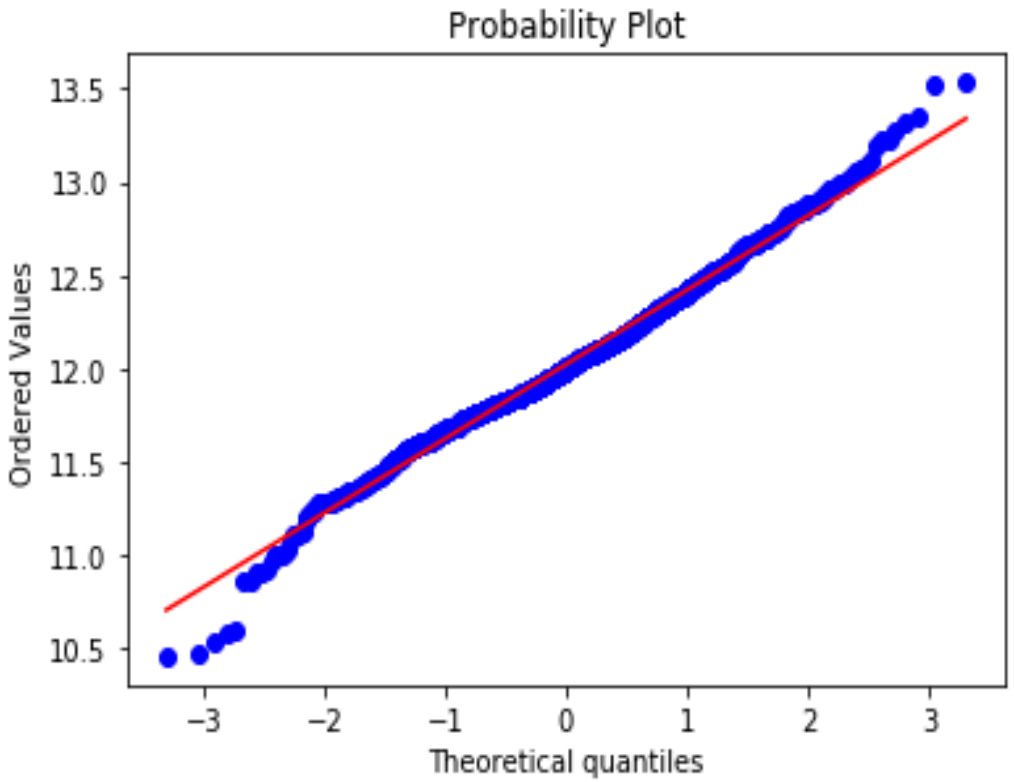
Assumptions for Linear Regression

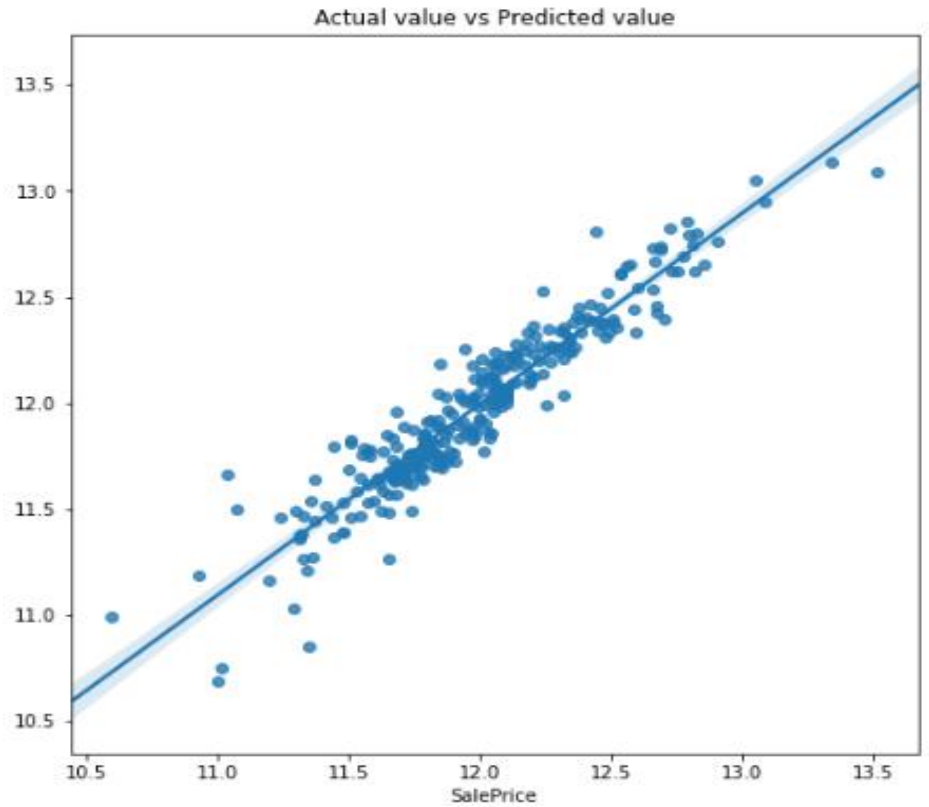
- **Independent features are not correlated - No multicollinearity**



Assumptions for Linear Regression

- **The error term must have constant variance – Homoscedasticity**





Comapring Models for Prediction

Linear Regression

	Bias	Variance	MAE	MSE	RMSE
0	0.006653	0.006653	0.102697	0.018963	0.137707

Random Forest

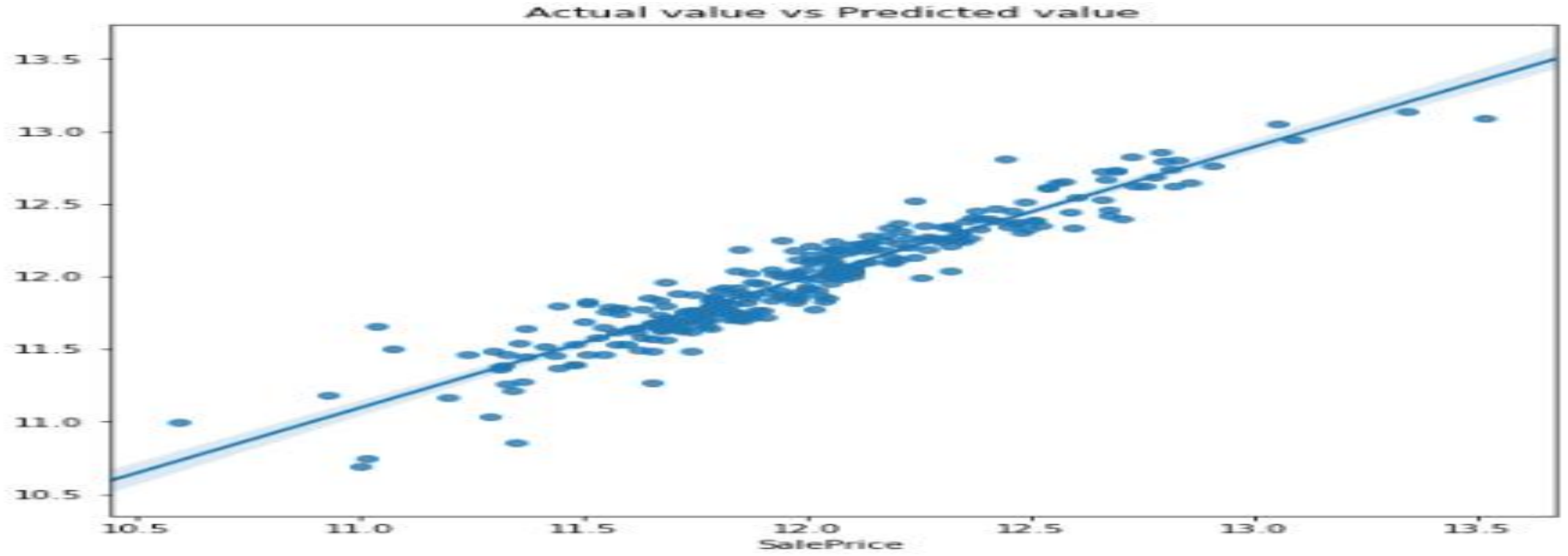
	Bias	Variance	MAE	MSE	RMSE
0	0.008144	0.008144	0.109968	0.024428	0.156295

Decision Tree

	Bias	Variance	MAE	MSE	RMSE
0	0.003795	0.003795	0.135125	0.034486	0.185705

Conclusion

Since errors along with Bias and Variance is low in Linear regression, so we will prefer Linear regression model



	Bias	Variance	MAE	MSE	RMSE
0	0.006653	0.006653	0.102697	0.018963	0.137707

THANK YOU