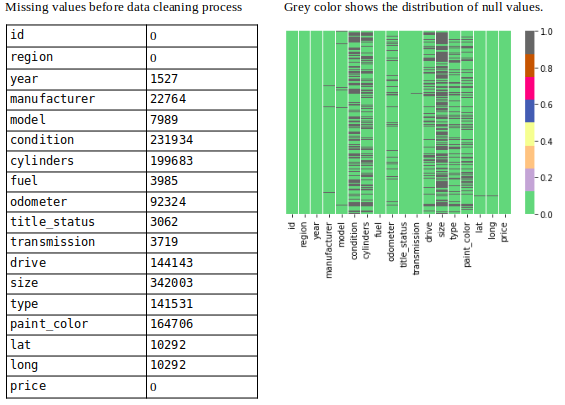
***Coding,Layout,Readability and Reusability***

**Data Cleaning:**

The first step is to remove irrelevant/useless features like ‘URL’, ’region\_url’, ’vin’, ’image\_url’, ’description’, ’county’, ’state’ from the dataset.

As a next step, check missing values for each feature.

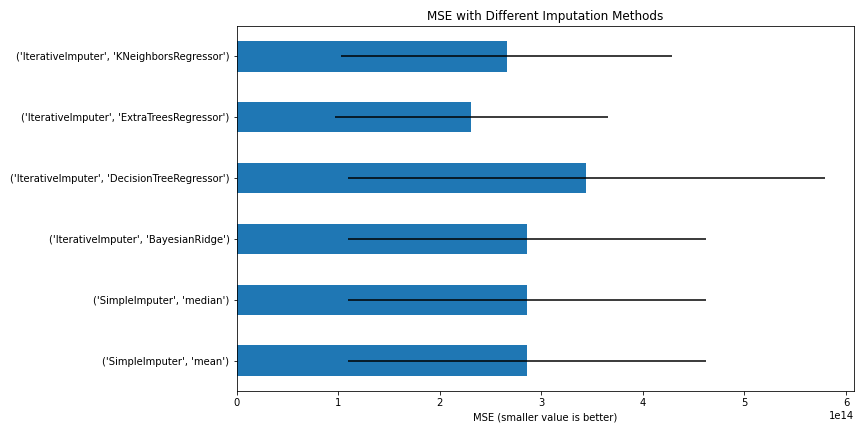


Showing missing values (Image By Panwar Abhash Anil)

Next, now missing values were filled with appropriate values by an appropriate method.

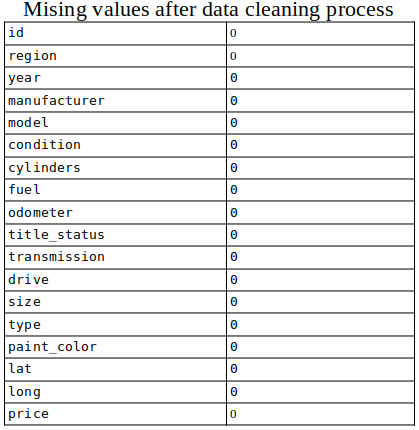
To fill the missing values, *[IterativeImputer](https://scikit-learn.org/stable/modules/generated/sklearn.impute.IterativeImputer.html" \t "_blank)* method is used and different estimators are implemented then calculated [*MSE*](https://en.wikipedia.org/wiki/Mean_squared_error) of each estimator using *[cross\_val\_score](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html" \t "_blank)*

1. Mean and Median
2. BayesianRidge Estimator
3. DecisionTreeRegressor Estimator
4. ExtraTreesRegressor Estimator
5. KNeighborsRegressor Estimator



MSE with Different Imputation Methods (Image By Panwar Abhash Anil)

From the above figure, we can conclude that the *ExtraTreesRegressor*estimator will be better for the imputation method to fill the missing value.



Missing values after filling (Image By [Panwar Abhash Anil](https://www.linkedin.com/in/abhash-panwar-85126976/))

At last, after dealing with missing values there zero null values.

**Outliers:**InterQuartile Range (IQR) method is used to remove the outliers from the data.

