1. **Data Set:** The data set is taken from ‘House Prices Prediction’ is taken from Kaggle. The data set has 81 columns with 1460 rows in it. All the variables explain about the features of the house such as Square footage, number of bedrooms and bathrooms, area of basement, garage capacity etc.
2. **Data Cleaning:**
3. Categorical Features:

* There are missing values in columns ‘MasVnrType’ and ‘Electrical’. I introduced ‘missing’ as another category for both the variables.
* Some of the values were coded ‘NA’ if a feature of the house was missing, I introduced a new class for such variables. For e.g. If a house doesn’t have a basement the value was coded as ‘NA’, I changed this to ‘No Bsmnt’.

1. Numerical Features:

* I used imputation method for numerical variables. Imputation is a method of taking mean of the values in a feature and assigning it missing values.