**Steps for linear regression**

1. Set Working Directory
2. Get WOrking Directory
3. Cleaning of Data
   1. Missing Value Imputation
      1. Mean, Median, Mode
      2. If normal replace with mean
      3. If skewed replace with median
      4. If categorical replace with mode
   2. Outlier Detection and Capping
      * Detect Outlier using boxplot
      * Cap it
   3. Dummy Variable

N category -> n-1 dummy

1. Check for Correlations of the data
2. Remove multicolinearity
   1. a= lm(col to predict ~. , data = data)
   2. vif(a)
   3. VIF> 10
   4. Remove Highest First, greater than 10
   5. Repeat till all col are less than 10
3. Summary(a) : Summary of model with all vif less than 10
4. Check for \* values
5. Remove variables which don't have \* values
   1. Step()
   2. Manually : Highest p value first
   3. Till all col are having \* values
6. We get final Model with all \*
7. Check for Assumption
8. Not Linear : Remove outliers in y
9. Check for new data do every variable have a \*
10. Then Check assumptions again
11. Calculate RMSE
12. Plot the graph