1. **Steps for Simple linear regression**
2. Import data
3. EDA (Exploratory Data Analysis)
4. Univariate (Summary, check Outlier & Missing value)
5. Bivariate (Scatter plot , Correlation)
6. View, Head functions
7. Data Pre-processing
8. Missing Value Treatment
9. Data conversion (Should be numeric)
10. Outlier Treatment
11. Again EDA
12. Data Partition
13. Training Data (should be 70%)
14. Testing Data (Should be 30%)
15. Check correlation on training data
16. If variable are correlated move forward
17. If variable are not correlated than this algorithm will not use
18. Create model
19. Check accuracy of the model (R^2 & Adj R^2)

* Value of R^2 and Adj R^2 should be between 70% to 90%
* If value of R^2 or Adj R^2 is more than 90%, data is over fit
* If value of R^2 or Adj R^2 is less than 70%, data is under fit

1. For removing over fitting and under fitting, use

* Data transformation method (log ,inverse etc)
* Data sampling method (Random, cluster, systematic, stratified) for data partition
* Add more variable (multiple linear regression)

1. Again create model, check the value of R^2 and Adj R^2.
2. P value should be less than α (0.05)
3. Will get model coefficient (β0, β1)
4. Validation (apply model(β0, β1) on testing data
5. Model is correct if

* Value of R^2 and Adj R^2 between 70% to 90% (if not again create model)

1. Assumptions
2. Linearity (between x and y , correlation or Scatter plot)
3. Normality
4. Constant variance
5. Model error (Auto correlation)
6. If assumptions are not satisfied than use different data transformation methods (Log, inverse etc.)
7. Prediction
8. Conclusion

**2. Steps for Multiple linear regression**

1. Import data
2. EDA (Exploratory Data Analysis)
3. Univariate (Summary, check Outlier & Missing value)
4. Bivariate (Scatter plot , Correlation)
5. View, Head functions
6. Data Pre-processing
7. Missing Value Treatment
8. Data conversion (Should be numeric)
9. Outlier Treatment
10. Again EDA
11. Data Partition
12. Training Data (should be 70%)
13. Testing Data (Should be 30%)
14. Check correlation on training data
15. If variable are correlated move forward
16. If variable are not correlated than this algorithm will not use
17. If co-linearity and multi co-linearity present in data, create model (Use variable selection method forward, backward & both, step function in R) and check VIF.
18. If VIF is more than 5% for any variable than remove that variable in descending order, VIF should be less than 5% for all variables.
19. Create Final model
20. Check accuracy of the model (R^2 & Adj R^2)

* Value of R^2 and Adj R^2 should be between 70% to 90%
* If value of R^2 or Adj R^2 is more than 90%, data is over fit
* If value of R^2 or Adj R^2 is less than 70%, data is under fit

1. For removing over fitting and under fitting, use data sampling method (Random, cluster, systematic, stratified) for data partition or data transformation and again create model, check the value of R^2 and Adj R^2.
2. P value should be less than α (0.05) for all variable otherwise remove the variable
3. Will get model coefficient (β0, β1 etc)
4. Validation (apply model(β0, β1 etc) on testing data
5. Model is correct if

* Value of R^2 and Adj R^2 between 70% to 90%

1. Assumptions
2. Linearity (between x and y , correlation or Scatter plot)
3. Normality
4. Constant variance
5. Model error (Auto correlation)
6. If assumptions are not satisfied than use different data transformation methods (Log, inverse etc.)
7. Prediction
8. Conclusion