

### ### Flow of building Linear Regression model

1. Identify the input and output variable from the dataset
2. Perform EDA to understand the Data (This step can be done after missing value imputation)
3. Perform Missing value treatment
4. Perform Outlier treatment on numerical data
5. Perform encoding on categorical column of data (one hot encoding with drop first)
6. Build the Linear regression model using OLS. See the Fstat pvalue and each feature pvalue
7. Check all the Assumptions:
  - 7a First Check the Multicollinearity Assumption using VIF and drop the variables one by one
  - 7b Check normality assumption for residue and if it is violated, do the transformation on output variable
  - 7c Check Linearity
  - 7d Check Autocorrelation
  - 7e Check Homoscedasticity
8. Build the model with columns remained after checking the assumption and compute Rsquare and RMSE
9. Perform Feature Engineering
  - 9a. Backward Elimination
  - 9b. Forward Selection
  - 9c. RFEFinalise the feature subset which producing the better result
10. Check the overfitting nature of the model
  - using train-test performance
  - using cross validation
11. If model overfits perform Regularization
  - using Ridge
  - using Lasso
  - using EnetTuning the model for the best alpha using gridsearchCV  
Finalise the regularized model which producing the better result in terms of bias variance trade-off
12. To explore whether Rsquare/ RMSE can be improved further, perform feature interaction
  - Feature interaction will increase the R2 but it also increase the number of useless feature
  - Perform feature selection (preferably use p value based backward elimination)
  - Build your model with the reduced interactive feature and you may get improved performance.
  - As we introduced new features, check the multicollinearity and other assumption once again before freezing the final model.