- check the description (overall and columns)
- check the null values
- treat the null values
  - either remove them (if they are less in number)
  - or go with mean,median or mode according to the distribution or need.

- go for distribution checking
- treat the distribution (use transformation techniuqes)
  - it can be right skewed or left
    - treat it (Rules)
      - if the distribution is low level skewed --> log transformation
      - if the distribution is moderate level skewed --> sqrt tranformation
      - if the distribution is severe/high level skewed --> inverse transformation

  - check the model by fitting all the data without doing any tranformation
  - and with tranformation , and check the accuracy in both the cases

- check and treat the outliers
  - either remove them (if they are less in number)
  - use some techniuqes (quartile method range, only remove top 1% quantile method)

- check the correlation and treat it if it is high
  - heatmap or corr()
  - treat - either you merge two columns or drop one non-important column
  - non-important --- check -- f_resgression() which will tell you the importance
  - PCA (optional for you right now)

- MODEL CREATION PART (MAIN IN SUPERVISED LEARNING)

- linear regression
  - check the testing accuracy
  - if it is low, you can go with regularisation methods and/or you can go with non linear models

- non linear models
  - xgboost, random forest, decision tree and so on

- implement hyper parameter tuning
- cross validation techniques
  - try out grid search cv / random search cv