

- 1) Complete FE for model ready data -
 - a. One-hot encoding for Linear model
 - b. Label encoding for KNN/DT
 - c. XG Boost/GBM model
- 2) Model ready data available
- 3) Train - test split
- 4) Scaling - Normalization v/s Standardization (Normally distribution)
 - a. LR & KNN - Mandatory
 - b. DT - Optional
- 5) Choose- y
 - a. LR - Log Y
 - b. DT/KNN - Normal
- 6) Accuracy metrics -
 - a. R-square/RMSE
 - b. RMSE/MSE/MAE
- 7) Build the model
 - a. Linear Regression -
 - i. Assumptions validation
 - 1) Multicollinearity
 - 2) Normal distribution of
 - b. Validate on the test data -
 - i. Accuracy
- 8) Hyperparameter tuning -
 - a. LR/Lasso/Ridge - Lambda optimization
 - b. KNN - K-nearest neighbors,
 - c. DT - Tree, Sample level parameters
- 9) Choose the best model
- 10) Interpretation of the model
 - a. LR - Need to tell the $y = mx + c$ (tell about the m, log y to be converted back to exp)
 - b. KNN/DT - Feature importance plot (explain the feature in detail - link the FI with EDA)
- 11) Business Implication of the model - Interpretation of the most optimum model and its implication on the business
 - a. How?
 - b. Opportunity - \$100M REV