

*Explaining the reasoning behind grading criteria.*

*1. Deliverable -1 (Exploratory data quality report reflecting the following) a. Univariate analysis i. Univariate analysis – data types and description of the independent attributes which should include (name, meaning, range of values observed, central values (mean and median), standard deviation and quartiles, analysis of the body of distributions / tails, missing values, outliers b. Multivariate analysis i. Bi-variate analysis between the predictor variables and between the predictor variables and target column. Comment on your findings in terms of their relationship and degree of relation if any. Presence of leverage points. Visualize the analysis using boxplots and pair plots, histograms or density curves. Select the most appropriate attributes c. Strategies to address the different data challenges such as data pollution, outliers and missing values*

*2. Deliverable -2 (Feature Engineering techniques) a. Identify opportunities (if any) to create a composite feature, drop a feature b. Decide on complexity of the model, should it be simple linear mode in terms of parameters or would a quadratic or higher degree help c. Explore for gaussians. If data is likely to be a mix of gaussians, explore individual clusters and present your findings in terms of the independent attributes and their suitability to predict strength*

*3. Deliverable -3 (create the model ) a. Obtain feature importance for the individual features using multiple methods and present your findings*

*4. Deliverable -4 (Tuning the model) a. Algorithms that you think will be suitable for this project b. Techniques employed to squeeze that extra performance out of the model without making it overfit or underfit c. Model performance range at 95% confidence level What will be the correct approach for doing the assignment? Following all the above listed steps as specified in the assignment with elaboration on each task as detailed as possible without deviating from end goal. Import libraries and read the dataset. Explore Data Analysis.*

*Find out how one variable related to other and distributions of data. State your insights. Build a linear regression model to predict the house prices Try to find out important features or create new features to improve the performance for your model. Use appropriate cross validations techniques to find out the best predictor parameters and best prediction. What most of the students have missed in the assignment?*

*Transform the features to common scale for unbiased comparison as there is variation in features data. Treatment of outliers and checking for missing values/Nans are missing. Applying Cross validation techniques with optimal parameter tuning (e.g. using GridSearchCV) by experimentation is missing. Highlighting Feature importance and selecting best features is missing. Experimentation of different approaches on the given problem is missing. One should try attempting to revisit feature building process when a model accuracy is low.*

*4. Common mistakes made by students (Wrong Assumptions) Feature engineering could have been better. Few of the redundant features could have been dropped but it's not done. Model Tuning and evaluation could have been better with clear justifiable insights. EDA without proper elaboration on insights are missing. Final summary insights are either missing or very brief with no proper conclusion.*

*5. Step by Step insights could have been better to justify the task behind each step in the execution flow.*