1. Data Preprocessing:
   1. *Handle any other data quality issues, such as outliers or inconsistencies.*
   2. Convert categorical variables into numerical representations if necessary.
   3. Remove clearly useless columns to clean up the dataset.
2. Feature Selection:
   1. Identify the relevant features using a RandomForestClassifier.
   2. Evaluate feature importances to select the most informative features for imputing missing values.
3. Curating the Data:
   1. Split the dataset into training and testing sets based on the important features identified by the random forest for the specific column.
   2. Ensure both sets contain missing values for a realistic assessment of the imputation models.
   3. This process removes any irrelevant or redundant features to reduce noise and improve efficiency.
4. Imputation Model Training:
   1. Train four different imputers using the curated training and testing data:
      1. SimpleImputer
      2. LinearImputer
      3. DeepImputer
      4. IterativeImputer
   2. For each imputer, use the training set for the specific target column to fit the model and the testing set to evaluate its initial performance.
5. Imputation Model Selection:
   1. Compare the performance of the imputers using mean\_squared\_error and pearson\_corr.
   2. Choose the best-performing imputer for the target column based on these evaluation metrics.
6. Final Imputation:
   1. Use the selected imputer to perform the final imputations on the entire dataset.
   2. Ensure the imputer is specifically curated to address the missing data in the target column.
7. Iteration and Refinement:
   1. If necessary, iterate and refine the imputation process by trying different techniques or adjusting hyperparameters.
   2. Incorporate domain knowledge to improve the quality of imputations.
   3. Re-evaluate the performance and make further adjustments as needed.