**Preprocessing Workflow:**

| **Abstract Workflow Node (Operation)** | **Input(s)** | **Output(s)** | **Implementation** |
| --- | --- | --- | --- |
| Count null data | dataframe | Null counts | Count null data |
| Delete column with zero data | Dataframe, threshold value | Dataframe | Deletes columns with zero data with threshold |
| Separate categorial numerical | Dataframe | Categorial columns, numerical columns | Separate categorical and numerical values |
| Drop column with zero threshold | Dataframe, threshold value | Dataframe | Perform drop columns with a threshold value |
| Apply 1 plus log transformation | Dataframe, columns to transform | Dataframe | Perform np.log1p function from numpy on the dataframe |

### **Analysis Workflow:**

| **Abstract Workflow Node (Operation)** | **Input(s)** | **Output(s)** | **Implementation** |
| --- | --- | --- | --- |
| Plot boxplot | Dataframe, x-axis column, y-axis column, output directory | Boxplot\_of\_{y\_column}\_by\_{x\_column}.png | Implementation of EDA among the columns |
| Plot heatmaps | Dataframe, output directory | Correlation\_Matrix\_Heatmap.png | How do the different columns correlate with the house's sale price? |

**Evaluation Workflow:**

| **Abstract Workflow Node (Operation)** | **Input(s)** | **Output(s)** | **Implementation** |
| --- | --- | --- | --- |
| Model evaluation | Model name, model to be evaluated, test features, true value, output file | Dictionary containing evaluation metrics | Evaluate MSE and R2-score of the models |
| Hyperparameter tuning | Models, parameter grids, X\_train, y\_train | Best model, best parameter | Perform hyperparameter tuning using GridSearchCV for multiple models |