Raw\_housing\_data

1. Classify numerical and categorical data (many variables are encoded such that they misled pandas to infer wrongly about their type)
2. Handle missing values (some appears as string such as “None”)
3. Encode missing value in variable “HOA” as -1 to preserve the distribution
4. Winsorize numerical data to deal with outliers
5. Visualizing relationships between numerical variables and housing “sold\_price” with boxplots and scatterplots