(2) Calculate the mean, Q1, Q2 (median), Q3, and IQR of the target and print these values. (4pts)

[20]:

(3) Remove outliers in the target variable using IQR. Calculate and print the upper and lower bounds of the target variable using IQR, and print the Ids of the outliers that fall outside this range. (3pts)

Calculate and print the Variance Inflation Factor (VIF) for the explanatory variables included in X1.

If the VIF is 10 or higher, it indicates a high degree of multicollinearity. The goal is to ensure that all explanatory variables have a VIF below 10. Explain which variables should be removed and provide justification for your choices.

Calculate the Pearson's correlation coefficient between the explanatory variables in X2 and the target, and select the top 5 explanatory variables that have the strongest linear relationship with the target. Then, plot the scatter plots between the target and these selected explanatory variables.

After training a linear regression model using X2, compute and display the estimated coefficients, standard errors, and t-values for all explanatory variables, including the intercept. Additionally, calculate the p-values for each variable based on a two-tailed significance test. Present the results in a table format

Perform an F-test to test the significance of the trained linear regression model and print a table summarizing the test results

Calculate and print the R-squared and adjusted R-squared values.

This time, you will train the linear regression model after applying a log transformation to the target. First, add 1 to the original target values, then take the natural log to create a new target variable. Afterward, plot a histogram (with 100 bins) and a boxplot for this new target variable. (3 pts)

Which has more outliers, the original target or the log-transformed target? And what do you think is the reason for this?