Explain any transformations or modifications you made to the dataset

**A**: First need to remove any column with NA value. Second, some of the columns look like ratio number but not, like year built. Zip code. Some are not very difficult to use like longitude and latitude. Some columns have the same value, like city names. All those columns will be removed.

Also different variable has different units so may need to do standarlize.

Create two variables; **one** that will **contain** the variables **Sale Price** and **Square Foot of Lot** (same variables used from previous assignment on simple regression) and **one that will contain Sale Price and several additional predictors of your choice**. Explain the basis for your additional predictor selections.

A:

Execute a summary() function on two variables defined in the previous step to compare the model results. What are the R2 and Adjusted R2 statistics? Explain what these results tell you about the overall model. Did the inclusion of the additional predictors help explain any large variations found in Sale Price?

A:

Considering the parameters of the multiple regression model you have created. What are the **standardized betas (** for each parameter and what do the values indicate?

https://statisticsglobe.com/extract-beta-coefficients-from-linear-regression-model-r

A:

Calculate the **confidence intervals** for the parameters in your model and explain what the results indicate.

A: use the coefplot to visualize it

Assess the improvement of the new model compared to your original model (simple regression model) by testing whether this change is significant by performing **an analysis of variance**.

Perform casewise diagnostics to identify outliers and/or influential cases, **storing each function's output in a dataframe assigned to a unique variable name.**

Calculate the **standardized residuals** using the appropriate command, specifying those that are +-2, storing the results of large residuals in a variable you create.

<https://www.statology.org/standardized-residuals-in-r/>

**standard\_res <- rstandard(model)**

Use the appropriate function to show the sum of large residuals.

https://www.statology.org/residual-sum-of-squares-in-r/

A screenshot of a computer

Description automatically generated

Which specific variables have large residuals (only cases that evaluate as TRUE)?

Investigate further by calculating the **leverage**, **cooks distance**, and **covariance rations**. Comment on all cases that are problematics.

Perform the necessary calculations to assess the **assumption of independence** and state if the condition is met or not. (use correlation)

Perform the necessary calculations to assess the **assumption of no multicollinearity** and state if the condition is met or not.

<https://sscc.wisc.edu/sscc/pubs/RegDiag-R/no-multicollinearity.html>

Visually check the assumptions related to the residuals using the plot() and hist() functions. Summarize what each graph is informing you of and if any anomalies are present.

Overall, is this regression model unbiased? If an unbiased regression model, what does this tell us about the sample vs. the entire population model?