# Real Estate

Frederick Jones

2023-11-20

```
# Load necessary library
library(readr)
```

```
#Quadratic term: Size squared (SizeSq = Size^2) to capture the non-linear effect of size on price.

#Dichotomous term: Swimming Pool (Pool), a binary variable (1 if the house has a pool, 0 otherwise).

#Dichotomous vs. Quantitative interaction term: Interaction between Pool and Age (Pool*Age) to see if t

# Load the dataset
data <- read_csv("/cloud/project/real_estate_price_data.csv")  # Replace with your actual file path
```

```
## Rows: 200 Columns: 6
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (6): Size, SizeSq, Age, Pool, PoolAgeInteraction, Price
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
data$SizeSq <- data$Size^2
data$Pool <- as.factor(data$Pool) # Converting Pool to a factor

# Building the multiple regression model
model <- lm(Price ~ Size + SizeSq + Age + Pool + Pool:Age, data=data)
```

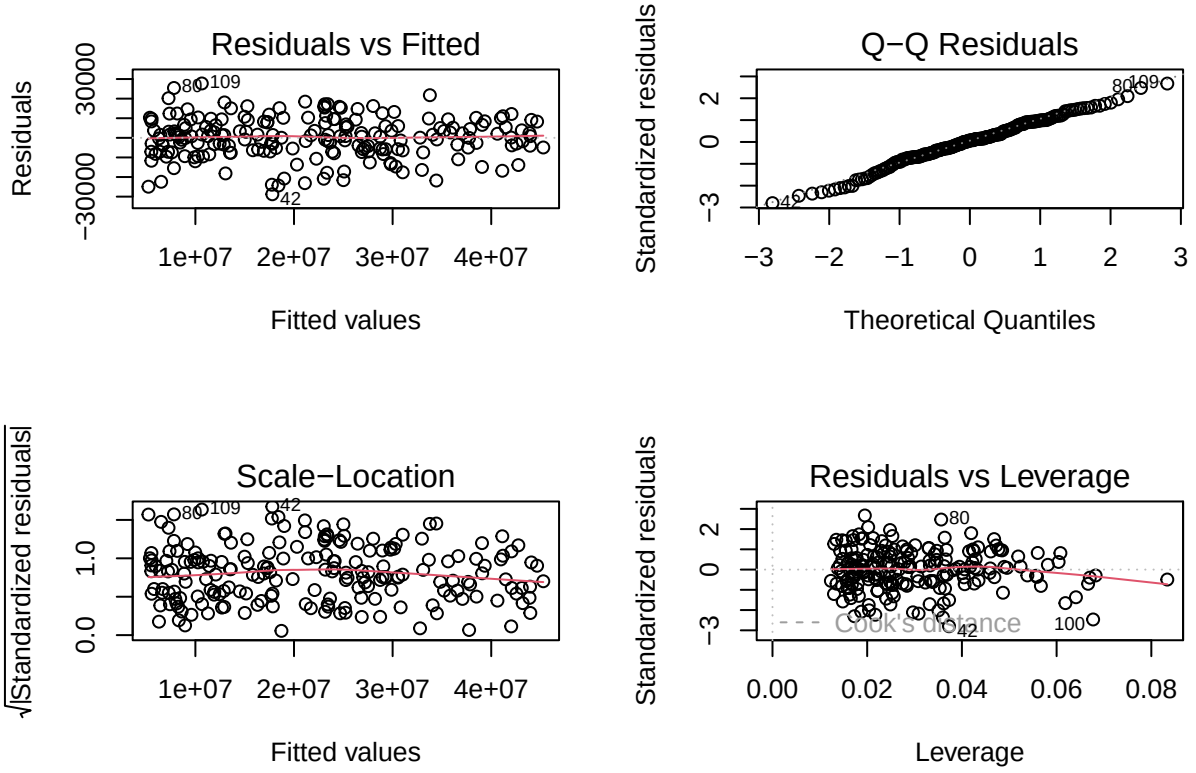
```
# Output the summary of the model
model_summary <- summary(model)
print(model_summary)
```

```
##
## Call:
## lm(formula = Price ~ Size + SizeSq + Age + Pool + Pool:Age, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28784.3  -6364.3    654.5   7613.1  27732.0
##
## Coefficients:
##               Estimate Std. Error  t value Pr(>|t|)
## (Intercept)  4.749e+04  1.000e+04    4.750 3.96e-06 ***
## Size         9.980e+01  9.999e+00    9.981  < 2e-16 ***
## SizeSq       5.000e+00  2.495e-03 2004.300  < 2e-16 ***
## Age         -1.577e+02  1.388e+02   -1.136    0.257
```

```
## Pool1         9.856e+03  3.841e+03    2.566    0.011 *
## Age:Pool1    -2.270e+02  2.040e+02   -1.113    0.267
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10480 on 194 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:      1
## F-statistic: 4.76e+07 on 5 and 194 DF,  p-value: < 2.2e-16
```

```r
# Interpretation of coefficients:
# Size: The effect of size on price, holding other factors constant.
# SizeSq: Captures the non-linear relationship of size with price.
# Age: Impact of age on price, controlling for other variables.
# Pool: Differential effect on price for houses with a pool vs. without.
# Pool:Age: Interaction effect showing how the impact of age on price differs for houses with a pool.

# Conducting Residual Analysis
par(mfrow=c(2,2))
plot(model)
```




```r
# The residual plots include:
# 1. Residuals vs Fitted: Checks for non-linearity and homoscedasticity.
# 2. Normal Q-Q: Checks for normality of residuals.
# 3. Scale-Location: Checks for homoscedasticity.
# 4. Residuals vs Leverage: Identifies influential cases.

# Evaluating Model Appropriateness
# - If residuals are randomly distributed with no clear patterns, and Q-Q plot shows a roughly straight
# - If there are clear patterns, trends, or deviations from normality in the residual plots, this may s
```

```
#Answer : The model can be said to be appropriate since the Q-Q plot shows a roughly straight line
```