model2.R

r2278750

2023-11-26

```r
# Multiple Linear Regression for Springbank Drive Data

# Clear Memory
rm(list = ls(all = TRUE))

# Import Data
dat <- read.csv("Springbank Drive Revised.csv", header = TRUE)

# Print column names on the screen
colnames(dat)
```

```
##  [1] "Property.."                      "Address"
##  [3] "Sales.Date"                      "HSETYPE"
##  [5] "One.and.a.Half.Storey"           "Two.Storey"
##  [7] "AGEYR"                           "LFA"
##  [9] "EXTAMEN"                         "Minor.Exterior.Amenities"
## [11] "Two.or.Three.Extra.Amenities"    "More.than.Three.Exterior.Amenities"
## [13] "EXTFINFACTOR"                    "Only.Brick"
## [15] "GAR"                             "Carport"
## [17] "One.Car.Garage"                  "Two.Car.Garage"
## [19] "STSCAPE"                         "Average.View"
## [21] "Good.View"                       "CENAIR"
## [23] "POOL"                            "INTCOND"
## [25] "Average.Interior.Condition"      "Good.Interior.Condition"
## [27] "Excellent.Interior.Condition"    "BSMTFINAREA"
## [29] "BI.AMEN.APPL"                    "LANESRD"
## [31] "TRAFCOUNT"                       "PRICE"
```

```r
# Variables for analysis
saleprice <- dat[,"PRICE"]  # Sale price of the property
lfa <- dat[,"LFA"]  # Area of Frontage
bsmtfinarea <- dat[,"BSMTFINAREA"]  # Basement Finished Area
ageyr <- dat[,"AGEYR"]  # Age of house in years
trafcount <- dat[,"TRAFCOUNT"]  # Amount of Traffic
onehalfstorey <- dat[,"One.and.a.Half.Storey"]  # Indicator for one and a half storey houses
twostorey <- dat[,"Two.Storey"]  # Indicator for two-storey houses
avgintcond <- dat[,"Average.Interior.Condition"]  # Indicator for houses with average interior conditio
goodintcond <- dat[,"Good.Interior.Condition"]  # Indicator for houses with good interior condition
excellentintcond <- dat[,"Excellent.Interior.Condition"]  # Indicator for houses with excellent interio
avgview <- dat[,"Average.View"]  # Indicator for houses with an average view
goodview <- dat[,"Good.View"]  # Indicator for houses with a good view
pool <- dat[,"POOL"]  # Indicator for presence of a pool
```

```r
# Generate a correlation matrix with the correct variables
cor(cbind(saleprice, lfa, bsmtfinarea, ageyr, trafcount, onehalfstorey,
          twostorey, avgintcond, goodintcond, excellentintcond,
          avgview, goodview, pool))
```

```
##                      saleprice          lfa   bsmtfinarea        ageyr   trafcount
## saleprice           1.00000000   0.40025782  0.2172499408 -0.134073554 -0.39711908
## lfa                 0.40025782   1.00000000 -0.0529680421  0.052779288 -0.15172499
## bsmtfinarea         0.21724994  -0.05296804  1.0000000000 -0.505577176 -0.12292764
## ageyr              -0.13407355   0.05277929 -0.5055771762  1.000000000  0.18254042
## trafcount          -0.39711908  -0.15172499 -0.1229276423  0.182540419  1.00000000
## onehalfstorey       0.04097100   0.30732441 -0.3380775920  0.344408666 -0.05950152
## twostorey          -0.06795999   0.13879021 -0.2743098231  0.294595253  0.31506234
## avgintcond         -0.21213447  -0.16718949  0.0004322962  0.005684107  0.15566992
## goodintcond         0.24754611   0.19742362  0.0642099698  0.033640944 -0.15084884
## excellentintcond    0.17764607  -0.03540225  0.1057009596 -0.272081408 -0.16136359
## avgview             0.03464353  -0.11178936  0.1749794978 -0.143459913  0.02624453
## goodview            0.23588527   0.11556279 -0.0619323057 -0.084446706 -0.36604648
## pool                0.15390020  -0.03808901  0.1664145984 -0.132126654 -0.27262159
##                   onehalfstorey    twostorey    avgintcond goodintcond
## saleprice            0.04097100  -0.06795999 -0.2121344692   0.24754611
## lfa                  0.30732441   0.13879021 -0.1671894883   0.19742362
## bsmtfinarea         -0.33807759  -0.27430982  0.0004322962   0.06420997
## ageyr                0.34440867   0.29459525  0.0056841075   0.03364094
## trafcount           -0.05950152   0.31506234  0.1556699220  -0.15084884
## onehalfstorey        1.00000000  -0.13124359  0.0640917781  -0.14683880
## twostorey           -0.13124359   1.00000000  0.0411909969  -0.13449056
## avgintcond           0.06409178   0.04119100  1.0000000000  -0.64672698
## goodintcond         -0.14683880  -0.13449056 -0.6467269782   1.00000000
## excellentintcond     0.10292773  -0.15399810 -0.3279504344  -0.25722086
## avgview             -0.15175738  -0.06639061  0.0737000447   0.03326739
## goodview             0.08319550  -0.09878983 -0.1580090351   0.06175986
## pool                -0.12139540  -0.16817499  0.0381001259   0.07681764
##                   excellentintcond      avgview     goodview         pool
## saleprice               0.17764607   0.03464353   0.23588527   0.15390020
## lfa                    -0.03540225  -0.11178936   0.11556279  -0.03808901
## bsmtfinarea             0.10570096   0.17497950  -0.06193231   0.16641460
## ageyr                  -0.27208141  -0.14345991  -0.08444671  -0.13212665
## trafcount              -0.16136359   0.02624453  -0.36604648  -0.27262159
## onehalfstorey           0.10292773  -0.15175738   0.08319550  -0.12139540
## twostorey              -0.15399810  -0.06639061  -0.09878983  -0.16817499
## avgintcond             -0.32795043   0.07370004  -0.15800904   0.03810013
## goodintcond            -0.25722086   0.03326739   0.06175986   0.07681764
## excellentintcond        1.00000000  -0.05623216   0.24265446  -0.05426380
## avgview                -0.05623216   1.00000000  -0.62225830   0.10965862
## goodview                0.24265446  -0.62225830   1.00000000  -0.00652692
## pool                   -0.05426380   0.10965862  -0.00652692   1.00000000
```

```r
# Multiple Linear Regression: Sales Price on multiple independent variables
mod_mlr <- lm(saleprice ~ lfa + bsmtfinarea + ageyr + trafcount + onehalfstorey +
              twostorey + avgintcond + goodintcond + excellentintcond + avgview +
              goodview + pool)

# Present Parameter Estimates, Coefficient of Determination, etc.
```

```r
summary(mod_mlr)
```

```
##
## Call:
## lm(formula = saleprice ~ lfa + bsmtfinarea + ageyr + trafcount +
##     onehalfstorey + twostorey + avgintcond + goodintcond + excellentintcond +
##     avgview + goodview + pool)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31867 -13663  -4346   8917 101178
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      80038.6857 19714.5233   4.060 0.000104 ***
## lfa                 36.1528    10.3326   3.499 0.000725 ***
## bsmtfinarea         18.8333     9.7160   1.938 0.055677 .
## ageyr              112.7235   190.1023   0.593 0.554677
## trafcount           -0.7341     0.3309  -2.219 0.029000 *
## onehalfstorey      731.2949 10509.3787   0.070 0.944677
## twostorey         8365.1886  7796.8599   1.073 0.286157
## avgintcond       10077.8006  8521.3101   1.183 0.240025
## goodintcond      18425.2365  9073.7116   2.031 0.045212 *
## excellentintcond 23717.9915 11244.1643   2.109 0.037659 *
## avgview           9056.6183  6329.4625   1.431 0.155893
## goodview         12500.5893  8130.6049   1.537 0.127647
## pool              5556.9444  7063.7966   0.787 0.433513
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22930 on 91 degrees of freedom
## Multiple R-squared:  0.3967, Adjusted R-squared:  0.3171
## F-statistic: 4.986 on 12 and 91 DF,  p-value: 2.69e-06
```

```r
# Extract standardized residuals and predicted values
standardized_residuals <- rstandard(mod_mlr)
predicted_saleprice <- predict(mod_mlr)

# Plot Predicted Sales Price vs Standardized Residuals
plot(predicted_saleprice, standardized_residuals, xlab = "Predicted Sale Price",
     ylab = "Standardized Residuals", main = "Predicted Sale Price vs Standardized Residuals")
```

Predicted Sale Price vs Standardized Residuals

Standardized Residuals

Predicted Sale Price