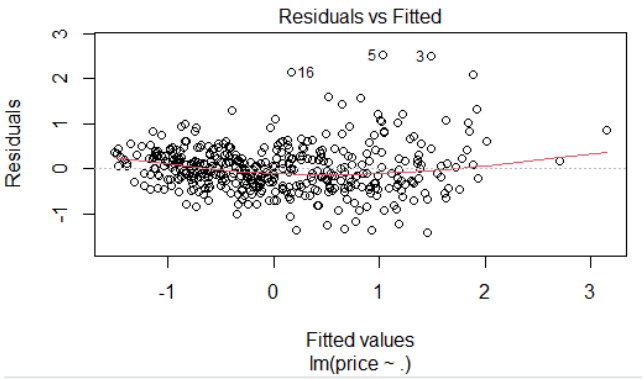
MICRO PROJECT




```
> summary(lm_model)

Call:
lm(formula = price ~ ., data = train_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.43425 -0.33007 -0.03193  0.27224  2.52018

Coefficients: (1 not defined because of singularities)
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -0.24018    0.06367  -3.773 0.000185 ***
area                             0.26829    0.02910   9.219  < 2e-16 ***
bedrooms                         0.02672    0.02965   0.901 0.368120
bathrooms                        0.26896    0.02890   9.307  < 2e-16 ***
stories                          0.23637    0.03118   7.582 2.19e-13 ***
parking                          0.13578    0.02808   4.836 1.86e-06 ***
mainroadno                      -0.22699    0.07846  -2.893 0.004014 **
mainroadyes                           NA         NA      NA       NA
guestroomyes                     0.18028    0.07141   2.525 0.011945 *
basementyes                      0.19691    0.06085   3.236 0.001307 **
hotwaterheatingyes               0.59405    0.12093   4.912 1.29e-06 ***
airconditioningyes               0.47084    0.06111   7.705 9.46e-14 ***
prefareayes                      0.32824    0.06500   5.050 6.59e-07 ***
`furnishingstatussemi-furnished` -0.03091    0.06528  -0.474 0.636086
furnishingstatusunfurnished      -0.21678    0.07081  -3.062 0.002342 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5368 on 422 degrees of freedom
Multiple R-squared:  0.7177,    Adjusted R-squared:  0.709
F-statistic: 82.53 on 13 and 422 DF,  p-value: < 2.2e-16

>
> # Make predictions on the test set
> predictions <- predict(lm_model, newdata = test_data)
>
> # Evaluate regression metrics
> cat("Metrics:\n")
Metrics:
> cat("MAE:", mae(predictions, test_data$price), "\n")
MAE: 0.5093771
> cat("RMSE:", rmse(predictions, test_data$price), "\n")
RMSE: 0.694099
```

```r
# Load necessary libraries
library(dplyr)
library(tidyr)
library(Metrics)
# Load the data
data <- read.csv("C:\\Users\\online\\Documents\\ML LAB\\Micro Project\\housing.csv")
# Preprocessing
# 4.1: Handling missing values
data <- na.omit(data)
# 4.2: Removing duplicate rows
data <- distinct(data)
# 4.3: Scaling numeric variables
numeric_columns <- c("price", "area", "bedrooms", "bathrooms", "stories", "parking")
data[numeric_columns] <- scale(data[numeric_columns])
# 4.4: Encoding categorical variables
categorical_columns <- c("mainroad", "guestroom",
 "basement", "hotwaterheating", "airconditioning", "prefarea", "furnishingstatus")
data <- model.matrix(~. - 1, data = data[, c(numeric_columns, categorical_columns)])
# Split the data into training and testing sets
set.seed(123)  # for reproducibility
train_indices <- sample(seq_len(nrow(data)), size = 0.8 * nrow(data))
train_data <- data[train_indices, ]
test_data <- data[-train_indices, ]
# Convert train_data matrix to data frame
train_data <- as.data.frame(train_data)
test_data <- as.data.frame(test_data)
# Fit Multiple Linear Regression model
lm_model <- lm(price ~ ., data = train_data)
# Print the Multiple Linear Regression model summary
summary(lm_model)
# Make predictions on the test set
predictions <- predict(lm_model, newdata = test_data)
# Evaluate regression metrics
cat("Metrics:\n")
cat("MAE:", mae(predictions, test_data$price), "\n")
cat("RMSE:", rmse(predictions, test_data$price), "\n")
plot(lm_model, which = 1)
```