TERMWORK - 01.

*. Problem Statement :-

Predict the price of a house by applying linear regression to using suitable dataset of real estate business.

*. THEORY:

Linear Regression is a statistical technique that aims to establish a relationship between a dependent variable and one or more independent variables.
By utilizing a dataset from the real estate business, which contains information about various houses & their corresponding prices, we can train a linear regression model. The model will learn the pattern and correlations between the independent variables and house prices, allowing it to make predictions on unseen data.

The practical involves the following steps:-

→ Data collection:- Gather a suitable dataset from the real estate domain.

→ Data processing:- Clean the dataset by handling missing values, removing outliers, & encoding categorical variables if necessary. Split the dataset into training set and a testing/validation set.

→ Model Training:- Apply linear regression to training set where the independent variables are used to predict the house prices.

→ Model Evaluation:- Evaluate the performance of the trained model using appropriate metrics, such as MSE, root MSE or R-squared. This step helps assess how well the model fits the training data & its ability to generalize the unseen data.

→ Predict house prices:- Apply the trained model to the testing/validation set to make predictions on house prices. Compare the predicted prices with actual prices to evaluate the model's accuracy.

→ Fine-tuning and improvement:- Analyze the results and consider fine-tuning the model by adjusting hyperparameters or exploring more advanced techniques, such as feature selection, regularization or ensemble methods, to enhance the prediction accuracy.

*. PROGRAM:

```
# load BostonHousing Data.
# Machine Learning Benchmark Problems.
library(mlbench)
# data manipulation library
install.packages("dplyr")
library(dplyr)
install.packages("ggplot2")
library(ggplot2)
install.packages("reshape2")
library(reshape2)
housing <- BostonHousing
str(housing)

# ggplot
housing %>%
ggplot(aes(x = medv)) +
stat_density() +
labs(x = "Median value($1000s)", y = "Density", title =
"Density Plot of Meadian Value House Price in Boston") +
theme_minimal()

# summary
summary(housing$medv)

#predicted v/s original
housing %>%
select(c(crim, rm, age, rad, tax, lstat, medv)) %>%
melt(id.vars = "medv") %>%
```

```
ggplot(aes(x = value, y = medv, colour = variable)) +
geom_point(alpha = 0.7) +
stat_smooth(aes(color = "black")) +
facet_wrap(~variable, scales = "free", ncol = 2) +
labs(x = "Variable Value", y = "Median House Price ($1000s)") +
theme_minimal()

#Set a seed of 123 and split your data into a train and
test set using a 75/25 split.
library("caret")
set.seed(123) # random no. generation
to.train <- createDataPartition(y = housing$medv, p = 0.75,
                                list = FALSE)
to.test <- createDataPartition(y = housing$medv, p = 0.25, list =
                               FALSE)

train <- housing[to.train, ]
test <- housing[to.test, ]

#fit a linear model
first_lm <- lm(medv ~ crim + rm + tax + lstat, data = train)

lm1_rsqu <- summary(first_lm)$r.squared
print(paste("First linar model has an r-squared value of",
      round(lm1_rsqu, 3), sep = " "))
## [1] "First linear model has an r-squared value of 0.6
# plot(first_lm)
```

```
#fix few problems
second-lm <- lm(log(medv) ~crim +rm +tax +lstat, data =train)

lm2-rsqu <- summary (second-lm) $r.squared
print(paste("our second linear model has an r-squared
value of", round(lm2-rsqu, 3), sep=" "))

abs(mean(second-lm $residuals))

#Create a dataframe of your predicted values & original ones
predicted <- predict(second-lm, newdata = test)
results <- data.frame(predicted = exp(predicted), original
                                  = test$medv)

#Plot this to visualize the performance of your model.
results %>% %>%
  ggplot(aes(x = predicted, y = original)) +
  geom-point() +
  stat-smooth() +
  labs(x = "Predicted values", y = "Original values", title =
    "Predicted vs. Original values") +
  theme_minimal()
```

* Conclusion:

In this teamwork we gained hands-on experience with linear regression, understand predictive capabi of model & learn how to apply it to real-world scenarios in the real estate business.