**QUESTION NO.1 OBSERVATIONS**

CRIME­­\_ RATE :

1. THE CRIME RATE VARIES WIDLEY WITH A RANGE OF 9.95
2. THE DISTRIBUTION IS WIDLEY POSITIVE SKEWED WITH A 0.021728

AGE:

1. DISTRIBUTION OF DATASET OF AGE IS NEGETAVILEY SKEWED
2. AGE OF PROPERTY IS RELATIVELY HIGH STANDARD DEVIATION.

INDUS:

1. DISTRIBUTION IS NEGATIVELY SKEWED
2. PROPRATION HAS LARGE STANDARD DEVIATION.

NOX:

1. NOX HAS MODERATE STANDARD DEVIATION
2. DISTRIBUTION HAS POSITIVELY SKEWED.

DISTANCE:

1. AVERAGE DISTANCE TO EMPLOYMENT IS AROUND 9.5 UNITS
2. DISTRIBUTION HAS A POSTIVIE SKEWE.

TAX:

1. TAX RATE WIDLEY IS WITH A RANGE IS 52.4
2. DISTRIBUTION HAS POSTIVE SLKEWE

PTRATIO:

1. PTRATIO IS RELATIVELY HIGH STANDARD DEVIATION
2. DISTRIBUTION IS NEGETAVILEY SKEWED

AVG-ROOM:

1. AVERAGE NUMBER OF ROOMS IN PROPERTIES WITH A RANGE 5.219
2. DISTRUBITON IS POSTIVE SKEWED

LSTAT:

1. PERCENTAGE OF LOWER STATUS OF POPULATION WITH A RANGE IS 36.24
2. DISTRIBUTION IS POSITVIELY SKEWED.

**QUESTION NO.2**

DISTRIBUTION OF DATA:

THE DATA IS GROUPED INTO VARIOS BINS OR INTERVALS . EACH BIN REPRESENTS A RANGE OF VALUES FOR A VARIABLE , AND THE FREQUENCY TELLS US HOW MANY DATA POINTS FALL WITHIN THAT RANGE.

DATA SUMMARY:

THE FREQUENCY DISTRIBUTION PROVIDES A SUMMARY OF THE DATA’S DISTRIBUTION MAKING IT EASIER TO UNDERSTAND THE DATA’S CHARCTERISTICS AND CENTRAL TENDENCIES.

1. There Are Two Data Points In The Bin Range Of 5.

2. There Are Four Data Points In The Bin Range Of Approximately 7.045.

3. There Are Fifteen Data Points In The Bin Range Of Approximately 9.091.

4. There Are Fourteen Data Points In The Bin Range Of Approximately 11.136.

5. There Are Twenty-Two Data Points In The Bin Range Of Approximately 13.182.

6. There Are Forty-Four Data Points In The Bin Range Of Approximately 15.227.

7. There Are Thirty-Two Data Points In The Bin Range Of Approximately 17.273.

8. There Are Fifty-Three Data Points In The Bin Range Of Approximately 19.318.

9. There Are Seventy Data Points In The Bin Range Of Approximately 21.364.

10. There Are Seventy-Seven Data Points In The Bin Range Of Approximately 23.409.

**QUESTION NO.3**

**OBSERVATION**

CRIME RATE:

STRONG POSITVIE COORELATIONWITH VARIABLE LIKE “TAX”(0.79) AND “LSTAT”(0.72).

AGE:

MODERATE POSITIVE CORRELATION BETWEEN “AGE” AND “TAX”(0.56)

INDUS:

THERE IS A WEEK NEGATIVE CORRELATION BETWEEN “INDUS” AND “CRIME RATE”(-0.11)

NOX:

THERE IS A WEEK POSTIVE RELATION BETWEEN “NOX”AND “TAX”(0.05).

DISTANCE:

MODERATE NEGATIVE COORELATION WITH VARIABLE LIKE “AGE”(-0.23)AND “NOX”(0.23)

TAX:

STRONG POSTIVE RELATION WITH VARIABLE LIKE “CRIME RATE”(0.79) AND “LSTATE”(0.65)

PTRATIO:

WEEK POSITIVE RELATION WITH VARIABLE LIKE “PTRATIO” “TAX”(0.07)

AVG ROOM:

WEEK POSITIVE RELETION WITH VARIABLE LIKE “AVG ROOM”AND “LSTATE”(0.49).

**QUESTION NO.4**

WE CAN IDENTIFY THE PAIRS WITH THE HIGHEST POSITIVE AND NEGATIVE COORELATION COEFFIECNTS.

3.POSTITIVE CORRELATED PAIRS:

TAX AND AVG\_PRICE;

DISTANCE AND NOX;

AGE AND CRIME\_RATE.

3.NEGATIVE COORELATED PAIRS:

LSTAT AND AVG\_ROOM

LATAT AND AVG\_PRICE

AGE AND INDUS

**QUESTION NO.5**

a). INTERPRETAION REGRESSION OF THE SUMMARY OUTPUT.

MULTIPLE R: THE MULTIPLE COORELATION COEFFIECENT ® IS 0.738.

THIS VALUE REPRESENTS THE STENGTH AND DIRECTION OF THE LINEAR RELATIONSHIP BETWEEN THE PREDICTOR VARIABLE AND THE ACUTAL VARIABLE.

R SQUARE : THE COEFFIECNT OF DETERMINA IS 0.544.IT REPRESENTS THE PROPORTION OF THE VARIANCE IN THE DEPENDENT VARIABLE THAT CAN BE EXPLAINED ,BY THE INDEPNDENT VARIABLE .

ADJUSTED R SQUARE:THIS IS THE MODIFIED VERSION OF R SQUARE . IT IS 0.543, WHICH IS VERY CLOSE TO R SQUARE.

STANDARD ERROR:THE STANDARD ERROR IS APPROXIMATELY 6.216, IT MESURES THE AVERAGE ERROR BETWEEN THE OBSERVED VALUES AND THE PREDICTED VALUES.

b). IS LSTATE VARIABLE SIGNIFICANT FOR THE ANALYSIS BASED ON YOUR MODEL?

YES, THE LSTATE VARIABLE APPEARS TO BE SIGNIFICANT FOR THE ANALYSIS BASED ON THE MODEL. THE COEFFICIENT FOR LSTAT IS -0.950 WITH A VERY LOW P-VALUE (5.081E-88).THE LOW P-VALUE INDICATES THAT THE LSTAT VARIABLE IS STARISTICALLY SIGNIFICANT IS PREDICTING THE DEPERTANT VARIABLE .

**QUESTION NO.6**

A)

(Y=M1X1+M2X2+M3X3.........+MNXN+C)

Y=5.094787984×7+(-0.642358334) ×20+(-1.35827812)

Y=21.45807639, Y=$21,450

THE COMPANY IS QUOTING $30,000

THE COMPANY IS OVERCHARGING THE CUSTOMER

B)

Yes, Is Model Performance Being Better Than The Previous Model Because The Adjusted R Square Value Is Greater In This Model. The Adjusted R Value Of Is Model Is 0.637124475. The Adjusted R Value Of The Previous Model.

**QUESTION NO.7**

ADJUSTED R SQUARE:

The Adjusted R-Squared Is Approximately 0.6883, Which Is Quite Similar To The Adjusted R-Squared Of The Previous Model. It Indicates That This Model Explains About 68.83% Of The Variability In AVG\_PRICE

Intercept (29.2413):

The intercept is the predicted value of AVG\_PRICE when all independent variables are set to zero. In this model, when all other independent variables are zero, the predicted AVG\_PRICE is approximately 29.2413 USD.

Coefficients:

The coefficients for each independent variable indicate the change in AVG\_PRICE for a one-unit change in that variable, holding all other variables constant. Let's analyze the significance of each independent variable:

The Adjusted R Value Of This Model Is 0.688298647

The Coefficient Value Of CRIME RATE IS 0.04872514

The Coefficient Value Of AGE Is 0.032770689

The Coefficient Value Of INDUS Is 0.130551399

The Coefficient Value Of NOX Is -10.3211828

The Coefficient Value Of DISTANCE Is 0.261093575

The Coefficient Value Of TAX Is -0.01440119

The Coefficient Value Of PTRATIO Is -1.074305348

The Coefficient Value Of AVG-PRICE Is 4.125409152

The Coefficient Value Of LSTAT Is -0.603486589

The Intercept Value Is 29.24131526

**QUESTION NO.8**

1. INTERPRATION OF THE OUTPUT OF THIS MODEL.

ADJUSTED R SQUARE:

THE ADJUSTED R SQUARE VALUE IS 0.6887,WHICH MEANS THAT APPROXIMATELY 68.87%OF THE VARIABLE IN AVG-PRICE IS EXPLAINED THE SIGNIFICANT INDEPENDENT VARIABLE IN THIS MODEL.

THIS VALUE IS SLIGHTELY LOWER THAN THE ADJUSTED R-SQUARED OF THE PREVIOUS MODEL (0.6887 VS0.6887).

COEFFICENT FOR SIGNIFFICANT INDEPENDENT VARIABLE:

* AGE:A ONE UNIT INCREASE IN AGE IS ASSOCIAATED WITH A INCREASE IN AVG-PRICE OF APPORIMATALY 0.033
* INDUS : A ONE UNIT INCREASE IN INDUS IS A ASSOCIATED WITH AN INCREASE IN AVG-PRICE OF APPROXIMATELY 0.131.
* NOX : A ONE UNIT INCREASE IN NOX IS ASSOCIATED WITH AN DECREASE IN AVG-PRICE OF APPROXIMATELY -10.273.
* DISTANCE : A ONE UNIT INCREASE IN DISTANCE IS ASSOCIATED WITH AN INCREASE IN AVG-PRICE OF APPROXIMATELY 0.262
* U
* PTRATIO: A ONE UNIT INCREASE IN PTRATIO IS ASSOCIATED WITH A DECREASE IN AVG-PRICE OF APPROXIMATELY -1.072.

1. COMPARING THE ADJUSTED R-SQUARE VALUE.

THE ADJUSTED R-SQUARE VALUE OF THE MODEL IS (0.6887)IS SLIGHTLY LOWER THAN THE ADJUSTE R-SQUARE VALUE OF THE PREVIOUS MODEL (0.6887 VS 0.6936). THEREFORE, THE PREVIOUS MODEL PERFORSM SLIGHTLY BETTER IN TERMS OF EXPLAINING THE VARIABLITY IN AVG-PRICE BASED ON THE ADJUSTED R-SQUARED VALUE.

1. Sorting the coefficient in ascending order:
2. NOX (-10.273) NEGATIVE.
3. LSTATE (-0.605) NEGATIVE.
4. PTRATIO (-1.072) NEGATIVE.
5. TAX (-0.0145) NEGATIVE .
6. AGE (0.033) POSITIVE.
7. INDUS (0.131) POSITIVE.
8. DISTANCE (0.262) POSITIVE.
9. AVE-ROOM (4.125) POSITIVE.
10. **WRITE THE REGRESSION EQUATION FROM THIS MODEL.**

THE EQUATION ALLOWS TO PREDICT AVG-PRICE BASED ON THE VALUES OF THE SIGNIFICANT INDEPENDENT VARIABLE IN YOUR MODEL.

Y=M1X1+M2X2+M3X3+………………+MnXn

**(M=COEFFICENT, N=VARIABLE)**