**BUISINESS REPORT**

**Q1:**

**APPROACH:**

1: Using data analysis tool Created Descriptive statistics for all the variables

2: In each summary statistics table, most significant measure is inferred and added below for reference.

**INFERENCES:**

**CRIME RATE:**

In the locality, The **Average crime rate is found out to be around 5 per capita.**

**AGE:**

It is found that **majority** of the house in that locality is of **AGE 100**, which indicates most properties in that locality are **old properties.**

**INDUS:**

It is found that the **industries are spread evenly** around the town with majority of the industry covering 18+ acres of locality.

**NOX:**

**Average NOX** in that locality is **0.55**(parts per million) and is evenly spread around.

**DISTANCE:**

**Most of the properties** are located at 24miles **from the highway.**

**TAX:**

The Property Tax rate is **evenly spread around 408/$10,000** and is not different for different type of property.

**PTRATIO:**

**Average PTRATIO IS 18** with **Maximum count as 22**, which is good.

**AVG\_ROOM:**

Most of the houses in the locality have on an **Average of 5 rooms.**

**LSTAT:**

In that locality, **Average LSTAT** is found out to be **12%.**

**AVG\_PRICE:**

Most of the properties Price are Quoted at/around **$50,000**(among 506 properties)

**Q2:**

**APPROACH:**

1: Avg\_Price column values are copied into separate worksheet first.

2: Created Histogram chart for that variable with bin size of 4.0

3: Calculated Mean, Median which can be used for inference of chart

**INFERENCE:**

From the Histogram plotted, we can clearly see that the tail of distribution is more on the right side.

The average price values are trailing towards right, which indicates **Positive skewness**.

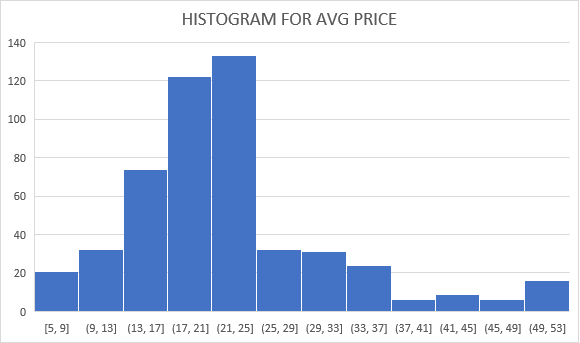
**Mean>Median**

We can say from the above histogram that Average price of properties that are valued above $25,000(approx.) diverse and less in count.

**TABLE:**



**HISTOGRAM CHART:**



**Q3:**

**APPROACH:**

1: Using Data analysis tool created covariance matrix for all the variables

2: Positive covariance: Both variable moves together in one direction

Negative covariance: One variable increase, the other variable decreases

3:Inferences are provided on high covariance pairs below based on above mentioned view

**COVARIANCE MATRIX:**



**INFERENCE:**

**POSITIVE COVARIANCE:**

(1)**Tax** INCREASES When the **Age of House** INCREASES

(2)**Tax** INCREASES as the **Distance from Highway** INCREASES

(3)**Tax** INCREASES When the **Proportion of Non-Retail Business acres per town** INCREASES

**NEGATIVE COVARIANCE:**

(1)**Average Price of House** DECREASES When **Tax** INCREASES

(2**) Average Price of House** DECREASES When **Age of House** INCREASES

(3)**Average Price of House** DECREASES When **LSTAT** INCREASES

**Q4:**

**APPROACH:**

1:Using Data analysis tool created Correlation matrix for all variables.

2: +1 Indicates Perfect Positive linear correlation between X and Y

0 Indicates no relation between X and Y

-1 Indicates Perfect Negative Linear Correlation between X and Y

3:Based on the above view, our inference is provided below with the table

**CORRELATION TABLE:**



**INFERENCE:**

**TOP 3 POSITIVELY CORRELATED PAIRS:**

|  |
| --- |
| *Tax, Distance* |
| *NOX, Indus* |
| *NOX, Age* |

**TOP 3 NEGATIVELY CORRELATED PAIRS:**

|  |
| --- |
| *AVG\_Price, LSTAT* |
| *LSTAT,AVG\_Room* |
| *AVG\_PRICE,PTRatio* |

**Q5:**

**APPROACH:**

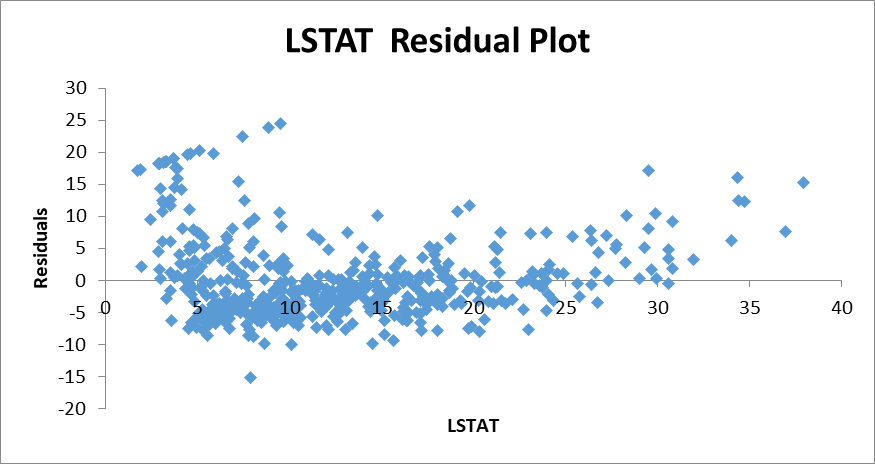
1:The LSTAT and Average price values are copied into separate worksheet

2:Using Data analysis tool, created Simple linear regression model and residual plot with Average Price as Y and LSTAT as X

**LINEAR REGRESSION EQUATION:**

**y=-0.95\*x1+34.55**

**GRAPH:**



**INFERENCE:**

**(a)**

**->VARIANCE EXPLAINED(R SQUARED):**

With LSTAT as Independent/Predictor variable, we can Determine 54% variance in the AVG\_Price of the properties.

**->COEFFICIENT VALUE:**

If the Independent variable(LSTAT) **INCREASES**, the AVG\_Price of the property **DECREASES BY 0.95%**

**->INTERCEPT:**

If the Value of **LSTAT becomes 0**,the AVG\_Price of the property will be **$34,553**

**->RESIDUAL PLOT:**

From the residual plot, it is clearly seen that the values are scattered randomly around Residual Line, **linear regression model is applicable**

------------------------------------------

**(b)IS LSTAT VARIABLE SIGNIFICANT FOR THE MODEL?**

The P-value of LSTAT variable is very less when compared to 0.05(alpha value).

And the **R-Square** value of this model is **around 54%** which is not bad.

Hence,LSTAT can be declared a **Significant Variable** when determining Price of property.

**Q6:**

**APPROACH:**

1:In addition to LSTAT,Average room is added as predictor to determine the average price of property

2:All the three values are copied into separate worksheet

3:Using Data analysis tool created regression model with one Y(Average price) and two X’s(LSTAT, Average room)

4:Inferences are shared below

**INFERENCE:**

1. **y=5.094787984\*x1-0.642358334\*x2-1.358272812** 

Given that average rooms are 7 and %Lower status of the population in that locality, Our Regression model **predicted the Average Price of the house can be $21,458(approx.)**

Comparing the Average Price predicted by our regression model($21,458) with the Price quoted by the company which is around $30,000 we can say that **the Company is overcharging for the property in that locality**

(b)



The performance of the Multilinear Regression Model is Better than our previous Simple Linear Regression model, Since the Adjusted R square value of Current model(0.64) **is greater than** the Adjusted R square value of previous model(0.54).

Hence,**In-addition to LSTAT variable,Avg room can also be added as predictor** which significantly increases the determination of the unknown variable(ie,Avg Price of house)

**Q7:**

**APPROACH:**

1:All the provided variables are taken as Predictors to determine the Average price of property in that locality

2:Using Data analysis tool created Multilinear regression model

3:This model is compared with previous model Adjusted R square, the coefficients, intercept, and significance of each variable with that of Average price is inferred below

NOTE: Correlation table is used to find the significance of each variable with Average price of property

**MULTILINEAR REGRESSION EQUATION:**

**y=0.0487\*x1+0.0328\*x2+0.1306\*x3-10.3212\*x4+0.2611\*x5-0.0144\*x6-1.0743\*x7+4.1254\*x8-0.6035\*x9+29.2413**

**INFERENCE:**

**ADJUSTED R SQUARE:**

In addition to the LSTAT and AVG\_Room variable, the addition of crime\_rate,age,Indus,NOX,Distance,Tax,PTRatio has increased the value of Adjusted R-Square from **0.63 to 0.68**,which indicates that addition of these variables improved our model’s prediction of property price a little bit.

**COEFFICIENTS:**

**CRIME RATE:-** Increase in crime rate Increases the price of property by **0.04%**

**AGE:-**Increase in AGE of property Increases price of property by **0.03%**

**INDUS:-**Increase in INDUS area Increases price of property by **0.13%**

**NOX:-**Increase in NOX value Decreases the price of property by **10%**

**DISTANCE:-**Increase in Distance from Highway Increases price of property by **0.26%**

**TAX:-**Increase in Property-Tax Decreases price of property by **0.01%**

**PTRATIO:-**Increase in PT\_Ratio Decreases price of property by **1%**

**AVG\_ROOM:-**Increase in AVG number of rooms Increases price of property by **4%**

**LSTAT:-**Increase in LSTAT of population Decreases price of property by **0.6%**

**INTERCEPT:**

When the values of all other Predictor variables are 0,the price of the property would be **$29,241**(approx.)

**SIGNIFICANCE:**

**CRIME RATE:**

Positive linear relation between crime rate and avg price of property, in which per crime rate increases ,avg price also increases by 4%

**AGE:**

Negative linear relation between crime rate and avg price of property, in which an increase in Age decreases avg price by 38%

**INDUS:**

Negative linear relation between crime rate and avg price of property, in which Indus increases avg price decreases by 48%

**NOX:**

Negative linear relation between crime rate and avg price of property, in which NOX increases avg price decreases by 43%

**TAX:**

Negative linear relation between crime rate and avg price of property, in which property-Tax increases avg price decreases by 47%

**DISTANCE:**

Negative linear relation between crime rate and avg price of property, in which Distance increases avg price decreases by 38%

**PTRATIO:**

Negative linear relation between crime rate and avg price of property, in which PTRatio increases avg price decreases by 54%

**AVG ROOM:**

Positive linear relation between crime rate and avg price of property, in which avg room increases avg price also increases by 70%

**LSTAT:**

Negative linear relation between crime rate and avg price of property, in which lstat increases avg price decreases by 74%

**Q8:**

**APPROACH:**

1:All the significant variables alone are taken into consideration for this model

NOTE: Significant variables are those whose P-Value is less than 0.05

2:Significant variables are:Age,Indus,NOX,Tax,Distance,PTRatio,AVG Room, LSTAT

3:Using Data analysis tool, Multilinear regression model has been created for those significant variables and the inferences are provided below

**MULTILINEAR REGRESSION EQUATION:**



**INFERENCE:**

(a)Interpretation of Model:

Predictors : age,Indus,NOX,Tax,Distance,PTRatio,AVG\_Room,

LSTAT

Prediction:Avg\_price of the property

Multiple R:83% OF Variance in Avg\_price can be explained by this model

R Square:70% of the price of property can be Determined with these predictors

P Value: All the values are less than 0.05 which indicates all are significant variables

Avg Error: 18%

Accuracy:81.50%

(b)Adjusted R Square(previous model):**69.83%**

Adjusted R Square(Current model):**69.87%**

Our current model **performs better** than previous model after the elimination of Crime Rate as predictor variable.

(c)After sort:

NOX : -10.27

PTRATIO : -1.0717

LSTAT :-0.60516

TAX :-0.01445

AGE :0.032935

INDUS : 0.13071

DISTANCE : 0.261506

AVG\_ROOM :4.125469

If the value of NOX is more in locality, then the Avg\_price of the properties will **decrease by 10 times**

(d)REGRESSION EQUATION:

**Y=0.0329\*X1+0.1307\*X2-10.2727\*X3-0.0145\*X4+0.2615\*X5-1.0717\*X6+4.1255\*X7-0.6052\*X8+29.4284**