# BUSINESS REPORT

## TERRO'S REAL ESTATE AGENCY

RAJESH LAMANI AUDITOR

**1.** The observation that we infer from the summary is the following

### CRIME RATE:

The average per capita crime rate is around 4.87 and the most occurred rate is around 3.43, the data is positively skewed and is platykurtic which means the observations are trailing towards the right and as it has negative kurtosis it shows that the observations are more concentrated and few outliers are there.

### AGE:

In this the average age is around 68.57 and higher half of the age is around 77.5 and the most frequent age turns out to be 100. The minimum age is 2.9 whereas the maximum is around 100 which is the most frequent.

### INDUS:

In case of proportion of non-retail business the average turns out to be 11.13%, the data is positively skewed and is platykurtic which means the observations are trailing towards the right and as it has negative kurtosis it shows that the observations are more concentrated and few outliers are there.

### NOX:

The average nitric oxide concentration turns out to be 0.55 part per 10 million and the most frequent turns out to be 18.1, the minimum concentration turns out to be 0.46 whereas the maximum turns out to be 27.74.

### DISTANCE:

The average distance of the property from the highway turns out to be 9.55 miles, and the higher 50% population has a distance of 5 miles from the highway, the minimum miles turns out to be 1 and maximum is around 24.

## TAX:

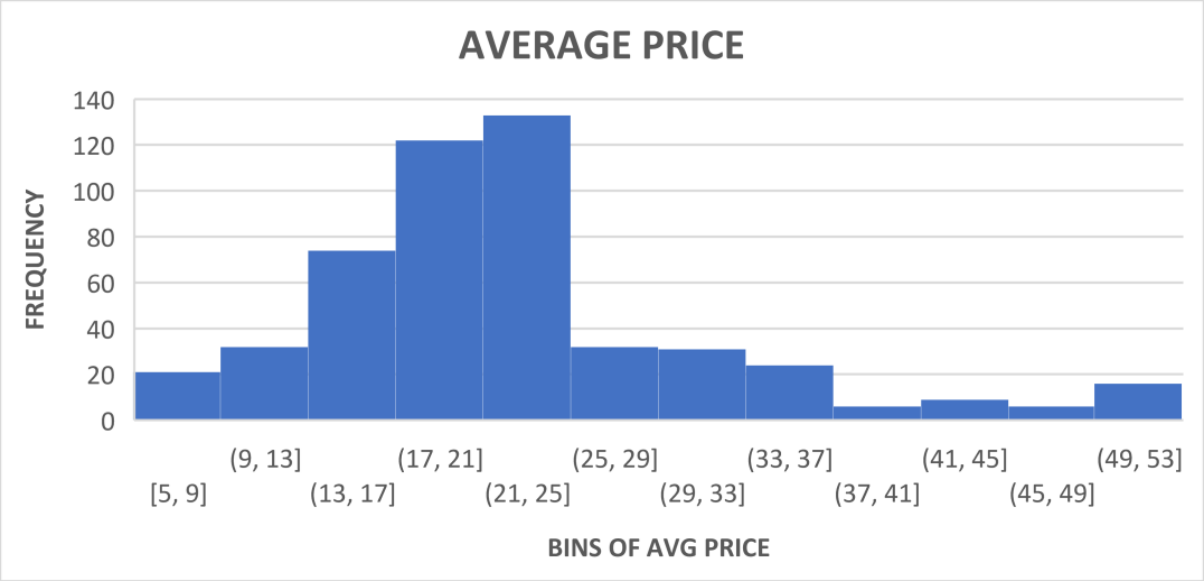
The average tax turns out to be 408.23 per $10,000, the minimum tax turns out to be around 187 and the maximum turns out to be 711.

## PTRATIO:

The average pupil to teacher ratio per town turns out to be 18.45 which means for every 1 teacher there are 18 students the higher 50% of the population has 19 students per 1 teacher, in this case the data is negatively skewed and platykurtic, which means the data is flat and trailing towards the left, and has less outliers and all the data is concentrated towards the between.

## AVG PRICE:

The average price of the houses turns out to be 22.53 ($1000) and the higher 50% of the price is around 21.2 ($1000 in this case the data is positively sewed and leptokurtic, which means that the data is trailing towards the right and has many outliers.




**2.**

From the above histogram data, we infer that the starting price range of the house turns out to be $5,000 the average price of the house is around 22.53 ($1000) but
the median is 21.2, indicating that the distribution is not symmetric, the data is

positively skewed and leptokurtic which means it has many outliers towards the higher price.

## 3.

The covariance matrix is a square matrix that measures the degree of linear relationship between two or more variables. A positive covariance indicates that two variables have positive relation with each other, while a negative covariance indicates vice -versa. A zero covariance means that there is no linear relationship between the variables.

The pair of variables with the highest positive covariance is

- age-tax
- distance-tax
- industry-tax

This means that these variables tend to increase and decrease together, suggesting a possible causal
relationship or common factor influencing them.

The pair of variables with the highest negative covariance is:

- tax-avg price
- age-avg price
- lstat-avg price

This means that these variables tend to move in opposite directions, implying an inverse relationship
or trade-off between them.

## 4.

**a)** Top 3 positively co-related pairs are

1) Tax & Distance
2) Nox & Indus
3) Nox & Age

b) Top 3 negatively co-related pairs are

1) Avg Price & Ptratio
2) Avg Room & LStat
3) Avg Price & Ptratio

## 5.

**a) R-squared**

This value tells us the proportion of the variance in the dependent variable (AVG_PRICE) that is predictable from the independent variable (LSTAT). If R-squared is greater than 50% it indicates a better fit to the model.

**INTERCEPT**

It tells us the avg price when LSTAT is zero. It's the point where regression line crosses the y axis

**RESIDUAL PLOT**

This plot shows the difference between the observed and predicted values (residuals) for each observation. If the residuals are randomly scattered around zero, it suggests that the model is a good fit. If the plot shows a linear pattern, it suggests that the model is missing some information.

**b)**

Significance on the basis of the LSTAT variable, The p-value for LSTAT in the regression summary tells us whether LSTAT is a significant predictor of AVG_PRICE. If the p-value is less than 0.05, we can conclude that LSTAT is a significant predictor.

**6.**

**a)**

Regression equation for two independent variable is

**Y = b + b1 * X1 + b2 * X2**

b = -1.35

b1 = 5.09

b2 = -0.64

Average Rooms = 7, L-STAT = 20

**AVG_PRICE** = -1.35 + 5.09*7 + (-0.64) *20 = 21.48

The predicted value according to the model is 21,480 which is less than 30,000 so in this case we can conclude that company is overcharging.

**b)**

Since the adjusted R square value for **this model is 0.6371** which is higher than **0.5432 in question 5**, in question 5 there was only one independent variable so if one more independent value is added and the adjusted r square is increasing so it shows the model in question 6 is better.

**7.**

In this case where we took all the variables to create the regression output the adjusted r square came out to be 0.68 whereas the r-square came out to be 0.69, this means that most of the predictor in this model are adding some values to the model

The intercept represents the estimated average price when all other predictors are zero. In this case, the intercept is statistically significant, suggesting that even when other predictors are zero, the average price is significantly different from zero.

Significance of each variable in respect to the AVG price is:

**CRIME_RATE**

Significance: Not significant (p-value = 0.5347)

Interpretation: The coefficient for CRIME_RATE is not statistically significant at the 0.05 significance level. This suggests that, based on the available data, there is not enough evidence to conclude that CRIME_RATE has a significant linear relationship with AVG_PRICE.

**AGE**

Significance: Significant (p-value = 0.0127)

Interpretation: The coefficient for AGE is statistically significant. A one-unit increase in AGE is associated with an increase in AVG_PRICE by 0.0328 units.

**INDUS**
Significance: Significant (p-value = 0.0391)
Interpretation: The coefficient for INDUS is statistically significant. A one-unit increase in INDUS is associated with an increase in AVG_PRICE by 0.1306 units.

**NOX**
Significance: Significant (p-value = 0.0083)
Interpretation: The coefficient for NOX is statistically significant. A one-unit increase in NOX is associated with a decrease in AVG_PRICE by 10.3212 units.

**DISTANCE**
Significance: Highly significant (p-value = 0.0001)
Interpretation: The coefficient for DISTANCE is highly statistically significant. A one-unit increase in DISTANCE is associated with an increase in AVG_PRICE by 0.2611 units.

**TAX**
Significance: Significant (p-value = 0.0003) Interpretation: The coefficient for TAX is statistically significant. A one-unit increase in TAX is associated
with a decrease in AVG_PRICE by 0.0144 units.

**PTRATIO**
Significance: Highly significant (p-value = 6.58642E−15)
Interpretation: The coefficient for PTRATIO is highly statistically significant. A one-unit increase in PTRATIO is associated with a decrease in AVG_PRICE by 1.0743 units.

**AVG_ROOM**
Significance: Highly significant (3.89287E-15)
Interpretation: The coefficient for AVG_ROOM is highly statistically significant. A one-unit increase in AVG_ROOM is associated with an increase in AVG_PRICE by 4.1254 units.

**LSTAT**
Significance: Highly significant (8.9107E-27)
Interpretation: The coefficient for LSTAT is highly statistically significant. A one-unit increase in LSTAT is
associated with a decrease in AVG_PRICE by 0.6035 units.

## 8.

**a)** This model is significant enough as the R square value is 0.6936, and the adjusted r square is also closer to the actual r square which shows that the predictor in this model are significant predictors.

**b)** The adjusted r square of the previous model was 0.688298647 whereas the current model adjusted r square is 0.688683682 which shows that the current model excluding crime rate is minutely better as compared to the previous model.

| | |
|---|---|
| NOX | -10.27270508 |
| PTRATIO | -1.071702473 |
| LSTAT | -0.605159282 |
| TAX | -0.014452345 |
| AGE | 0.03293496 |
| INDUS | 0.130710007 |

DISTANCE 0.261506423
AVG_ROOM 4.125468959

**c)** From the table data we can see if we sort the coefficient value in an ascending order, we got to know that NOX is the most negative coefficient, so it will have a negative impact on the price. if the NOX concentration increases the average price of the house will decrease.

**d)** The regression equation of this model will be
AVERAGE PRICE = 29.42 + (0.032*AGE) + (0.130 * INDUS) + (-10.27 * NOX) + (0.261 * DISTANCE) +(0.0144 * TAX) + (-1.0717 * PTRATIO) + (4.125 * AVG_ROOM**)** + (-0.605*LSTAT)