3/16/2023

Terro’s Real Estate

BUSINESS REPORT

DINESH M
[COMPANY NAME]

SUMMARY

Terro's real-estate is an agency that estimates the pricing of houses in a certain locality. The pricing is concluded based on different features of a property. This also helps them in identifying the business value of a property. To do this activity the company employs an "Auditor", who studies various geographic features of a property like pollution level (NOX), crime rate, education facilities (pupil to teacher ratio), connectivity (distance from highway), etc. This helps in determining the price of a property. The agency has provided a dataset of 506 houses in Boston.

Data Dictionary

Crime Rate	per capita crime rate by town
Industry	proportion of non-retail business acres per town (in percentage terms)
NOX	nitric oxides concentration (parts per 10 million)
Average Rooms	average number of rooms per house
Age	proportion of houses built prior to 1940 (in percentage terms)
Distance	distance from highway (in miles)
Tax	full-value property-tax rate per $10,000
PT ratio	pupil-teacher ratio by town
LSTAT	% lower status of the population
Average Price	Average value of houses in $1000's

TABLE OF CONTENT

INTRODUCTION

Terro's real-estate is an agency that estimates the pricing of houses in a certain locality. From given data formed a regression model by using data analysis tool pak for each of the variables to analyse the intercept, coefficient, variance and residual plot and also made some of the methods to find the changes in average price of house.

1) Summary Statistics for Each Variable

Crime rate:

The Skewness is 0.021 its close to zero so it shows that symmetric distribution

The kurtosis is -1.18 is shorter peak and thicker tail than the normal distribution

Min and max is 0.04 – 9.99 falls between this value

Age:

The skewness is –0.598 its left skewed data

The kurtosis is -0.967 is shorter peak and thicker tail than the normal distribution

Min and max is 2.9 – 100 falls between this value

Indus:

The skewness is 0.295 its close to zero so it shows that symmetric distribution

The kurtosis is -1.23 is shorter peak and thicker tail than the normal distribution

Min and Max is 0.46 - 27.74 falls between this value

NOX:

The skewness is 0.729 its right skewed data

The kurtosis is -0.06 its close to zero the tails are consistent with normal distribution

Min and Max is 0.385 – 0.871 falls between this value

Distance:

The skewness is 1.004 its right skewed data

The kurtosis is -0.86 is shorter peak and thicker tail than the normal distribution

Min and Max is 1 – 24 fall between this data

PTRATIO:

The skewness is -0.802 its left skewed data

The kurtosis is -0.28 its close to zero the tails are consistent with normal distribution

Min and Max is 12.6 – 22 falls between this data

Avg room:

The skewness is 0.403 its close to zero so it shows that symmetric distribution

The kurtosis is 1.89 its higher peak and thicker tail than the normal distribution

Min and Max is 3.561 – 8.78 falls down between this data

LSTAT:

The skewness is 0.906 its right skewed data

The kurtosis is 0.493 its close to zero the tails are consistent with normal distribution

Min and Max is 1.73 – 37.97 falls down between this data

Avg Price:

The Skewness is 1.108 its right skewed data

The kurtosis is 1.495 its higher peak and thicker tail than the normal distribution

Min and Max is 5 – 50 fall down between this data

Tax:

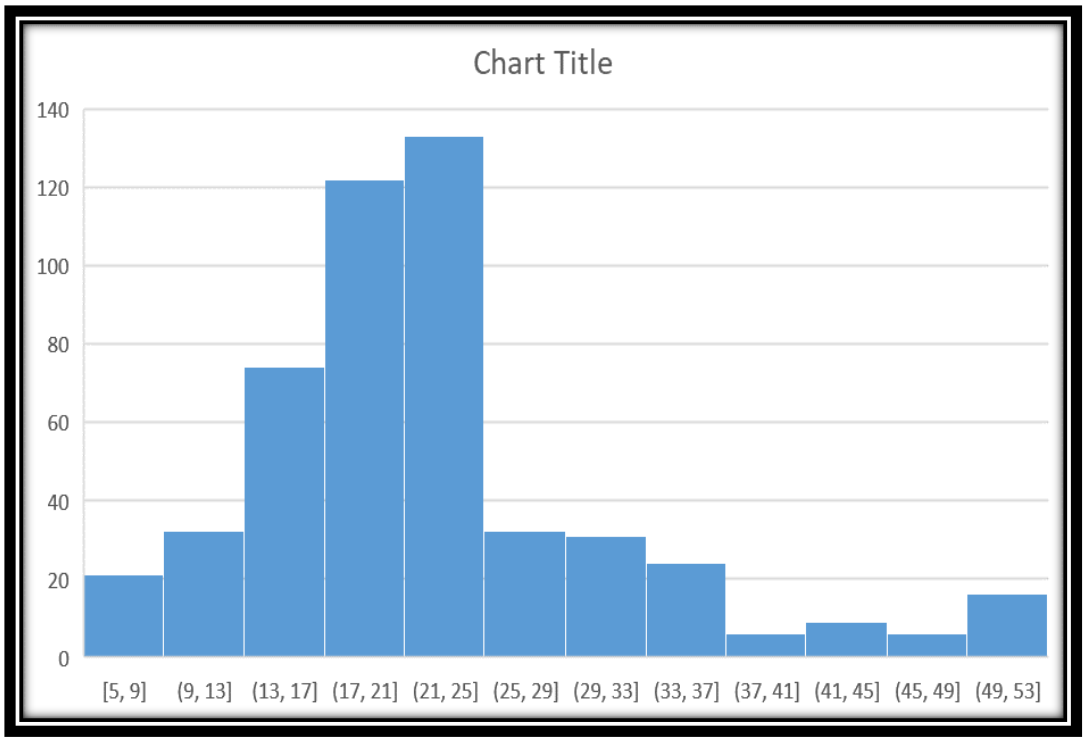
The skewness is 0.669 its right skewed data

The kurtosis is -1.14 is shorter peak and thicker tail than the normal distribution

Min and Max is 187 – 711 fall down between this data

2) Plot histogram for Average Price

Plotting the histogram of average price of house range between 21 to 25 has the highest count (133) then the other range.



3) Covariance matrix

Computing the covariance matrix for a given data. The diagonal elements of covariance matrix is a variance of individual variance and its should be always in positive value. In this all the diagonal elements are positive. The remaining off diagonal elements is covariance of paired variables. It shows how the two variables are vary. Its in positive value so it is good for interpreting its magnitude. It denotes direct connection.

4) Correlation matrix

Correlation shows the relationship between them and how they related

a) Top 3 positive pairs

The top 3 positive pairs are

i) Tax and Distance (0.91)
ii) NOX and Indus (0.76)
iii) NOX and Age (0.73)

b) Top 3 negative Pairs

The top 3 negative pairs are

i) LSTAT and Avg Price (-0.73)
ii) LSTAT and Avg Room (-0.61)
iii) PTRATIO and Avg Price (-0.50)

5) Regression Model Average Price (Y) and LSTAT (X

a) The coefficient of LSTAT is -0.95005 it is negative relationship between X and Y so its not a good model, LSTAT Residual plot are Homoskedastic because the trendline of linear line is in centre.
b) The LSTAT p-value is 0% so it is significant.

6) Regression model LSTAT (X1), Average Room (X2) and Average Price (Y)

a) The average price value is 21.45 for 7 rooms and value of 20 for LSTAT so the company is overcharging the value of house for this locality.

b) In Question 5 adjusted R value is 0.5432. In Question 6 adjusted R value is 0.6371. Question 6 has the higher value of adjusted R value than Question 5.

7) Regression model for all variables

Adjusted R square value is 0.688299

Coefficient values

Crime rate = 0.048725

Age = 0.032771

Indus = 0.130551

NOX = -10.3212

Distance = 0.261094

Tax = -0.0144

PTRATIO = -1.07431

Avg Room = 4.125409

LSTAT = -0.60349

	Crime rate	Age	Indus	Nox	Distance	Tax	PTratio	Avg room	LSTAT
P value	53%	1%	4%	1%	0%	0%	0%	0%	0%

The individual P-value of crime rate is greater than alpha value (5%) so we can remove the crime rate.

8) Significance value for all variables

	Crime rate	Age	Indus	Nox	Distance	Tax	PTratio	Avg room	LSTAT
P value	53%	1%	4%	1%	0%	0%	0%	0%	0%

The individual P-value of crime rate is greater than alpha value (5%) so we can remove the crime rate.

Conclusion:

According to given data made we analysis the regression model with reputed data and also predicted the range, min , max, mean, median values of all the given data. Its shows that the terro's real estate agency can understand their house prices, crime rate range and tax of the property.