

LINEAR REGRESSION ASSIGNMENT QUIZ

How many employees having more than 5 years experience are earning more than 60000?

Marked Answer :

14

Correct Answer :

14

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

How many employees are earning between 50000-80000?

Marked Answer :

12

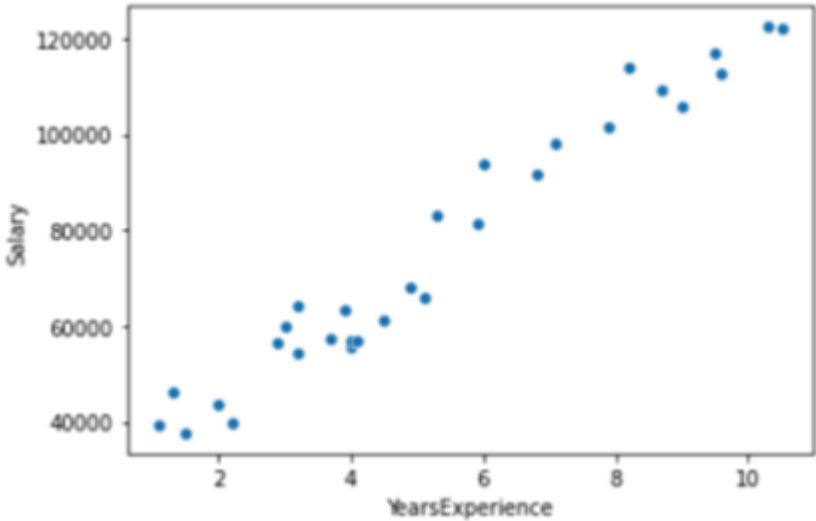
Correct Answer :

12

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The scatter plot in the following image shows the relationship between the “YearsExperience” and “Salary” columns. What possible inferences can be drawn from the plot?



Marked Answer :

The plot shows a positive correlation between the ‘YearsExperience” and “Salary” column.

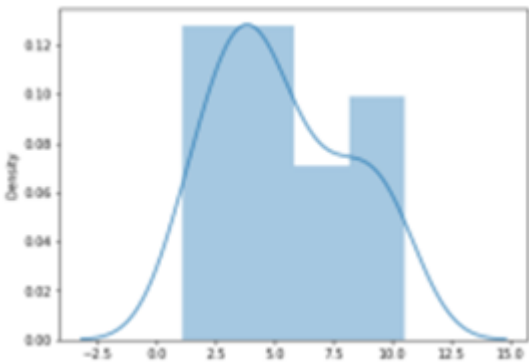
Correct Answer :

The plot shows a positive correlation between the ‘YearsExperience” and “Salary” column.

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The distribution plot of the column “YearsExperience” is shown in the image below, what possible inferences can be drawn from the plot.



Marked Answer :

“YearsExperience” data is normally distributed.

Correct Answer :

“YearsExperience” data is positively skewed.

MARKS OBTAINED ✗ 0

TOTAL MARKS : 1

What all inferences can be drawn from the table shown below:


	YearsExperience	Salary
count	30.000000	30.000000
mean	5.313333	76003.000000
std	2.837888	27414.429785
min	1.100000	37731.000000
25%	3.200000	56720.750000
50%	4.700000	65237.000000
75%	7.700000	100544.750000
max	10.500000	122391.000000

Marked Answer :

The range of the “YearsExperience” and “Salary” data is (9.4 , 84660)

Correct Answer :

The range of the “YearsExperience” and “Salary” data is (9.4 , 84660)

MARKS OBTAINED  1 TOTAL MARKS : 1


To split the dataset into training and testing data, if we use the following code. X = data[‘YearsExperience’]
y = data[‘Salary’]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0) What does it mean when we write the test size as 0.2?

Marked Answer :

The training data will consist of 80% of the samples from the total population.

Correct Answer :

The training data will consist of 80% of the samples from the total population.

MARKS OBTAINED  1 TOTAL MARKS : 1


In the above example code, we have taken the random state as 0, if we change the random state as 42, what does it mean for our training and testing data?

Marked Answer :

The random state does not have any effect on the shape of the data.

Correct Answer :

The random state does not have any effect on the shape of the data.

MARKS OBTAINED  1 TOTAL MARKS : 1


If the r2 score calculated in the above example is 0.98 , change the sample size of the training and testing set in the ratio 60:40, and build a linear regression model again. After plotting the best fit line on the test data, calculate the r2_score for the new model.

Marked Answer :

0.95-0.97

Correct Answer :

0.95-0.97

MARKS OBTAINED  1 TOTAL MARKS : 1


If while fitting the model with training and testing data, you get the following error ValueError: Expected 2D array, got 1D array instead: What could be the issue with the data, and how can you solve it?

Marked Answer :

Reshape the data to a two dimensional array

Correct Answer :

Reshape the data to a two dimensional array

MARKS OBTAINED  1 TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

How many houses have a waterfront?

Marked Answer :

163

Correct Answer :

163

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

How many houses have 2 floors?

Marked Answer :

8241

Correct Answer :

8241

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

How many houses built before 1960 have a waterfront?

Marked Answer :

80

Correct Answer :

80

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

What is the price of the most expensive house having more than 4 bathrooms?

Marked Answer :

7700000

Correct Answer :

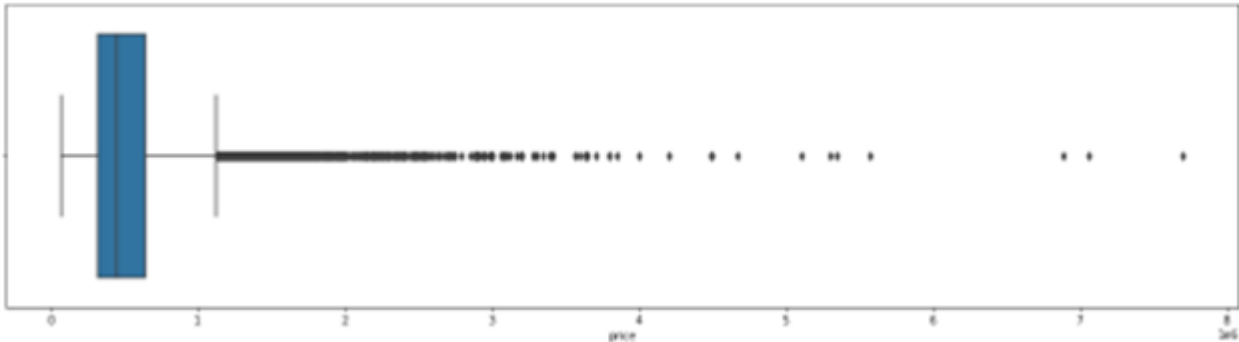
7700000

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

The image shown below shows the boxplot of the price column from the housing dataset. What inferences can you make from the plot?



Marked Answer :

There is a presence of outliers in the price data.

Correct Answer :

There is a presence of outliers in the price data.

MARKS OBTAINED ✓ 1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.


For instance, if the ‘price’ column consists of outliers, how can you make the data clean and remove the redundancies?

Marked Answer :

Calculate the IQR range and drop the values outside the range.

Correct Answer :

Calculate the IQR range and drop the values outside the range.

MARKS OBTAINED  1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.


What are the various parameters that can be used to determine the dependent variables in the housing data to determine the price of the house?

Marked Answer :

Correlation coefficients

Correct Answer :

Correlation coefficients

MARKS OBTAINED  1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.

If we get the r2 score as 0.38, what inferences can we make about the model and its efficiency?

Marked Answer :

Low difference between observed and fitted values.

Correct Answer :

High difference between observed and fitted values.

MARKS OBTAINED  0

TOTAL MARKS : 1


What will be the appropriate metrics for performance evaluation of Linear Regression model?

Marked Answer :

r2_score

Correct Answer :

r2_score

MARKS OBTAINED  1

TOTAL MARKS : 1

The exercise after this contains questions that are based on the housing dataset.


If the Variance Inflation Factor value for a feature is considerably higher than the other features, what can we say about that column/feature?

Marked Answer :

High multicollinearity

Correct Answer :

High multicollinearity

MARKS OBTAINED  1

TOTAL MARKS : 1

Total Marks

17 / 19