

Sales Model- Report

Parameters Used:

- `!pip install`: installs the required packages onto the local system.
- `import _____ as _____`: imports the required libraries into the file which can be called with the alias specified after 'as'.
- `read_csv()`: pandas command which reads the dataset from the link specified in the brackets.
- `plt.scatter(x,y)`: plot a graph of the data using matplotlib. Each individual data point is marked with a dot.
- `plt.plot(x,y)`: plots a line graph using matplotlib. Draws a line passing through the data points.
- `plt.xlabel()/plt.ylabel()`: adds a label to the axes of the graph. The label is specified in the brackets.
- `train_test_split`: it is a function in the model selection module of scikit learn library. This function splits the data into 4 arrays. It splits the dataset into training and testing subsets.
- `np.array().reshape(-1,1)`: reshapes the array into 1 column. '-1' means unknown dimension, which is used if you want numpy to automatically figure out the row size.
- `lr.predict()`: predicts the dependent variable data using linear regression model.

R-2 score:

- R-squared is commonly used to evaluate the accuracy of linear regression models.
- It is a value between 0 and 1. The closer the value is to 1, the more accurately it predicts the data.
- It is calculated as the proportion of the total sum of squares (RSS) the total sum of squares (TSS), which represents the total variability in the dependent variable.
- Mathematically it is represented as:

$$r^2 = 1 - (rss/tss)$$

rss is the sum of the squared differences between the actual values and the predicted values.

tss is the sum of the squared differences between the actual values and the mean of the dependent variable.