**Implementing Linear Regression in Python**

**LINEARREGRESSION**  
It is the basic and commonly used type for predictive analysis. It is a statistical approach to modelling the relationship between a dependent variable and a given set of independent variables.

**These are of two types:**

1:Simple linear Regression

2:Multiple Linear Regression

**Multiple linear regression** can model the relationship between two or more features and response by fitting a linear equation to observed data.

**STEP 1: IMPORTING PACKAGES AND CLASSES**

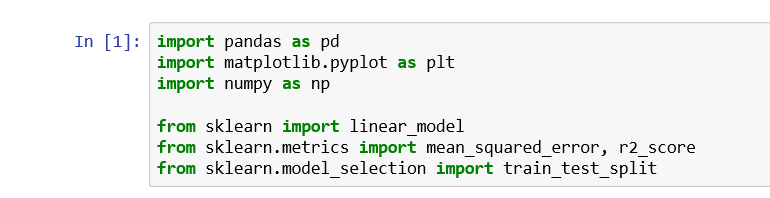
The first step is to import the package numpy and the class LinearRegression from sklearn.linear\_model:

**import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
**import** numpy **as** np  
  
**from** sklearn **import** linear\_model  
**from** sklearn.metrics **import** mean\_squared\_error, r2\_score  
**from** sklearn.model\_selection **import** train\_test\_split

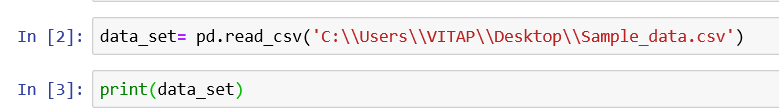
**NumPy** is the fundamental package for scientific computing.

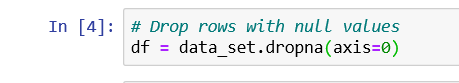
The **scikit-learn** is a widely used python library for machine learning, built on top of NumPy and some other packages.

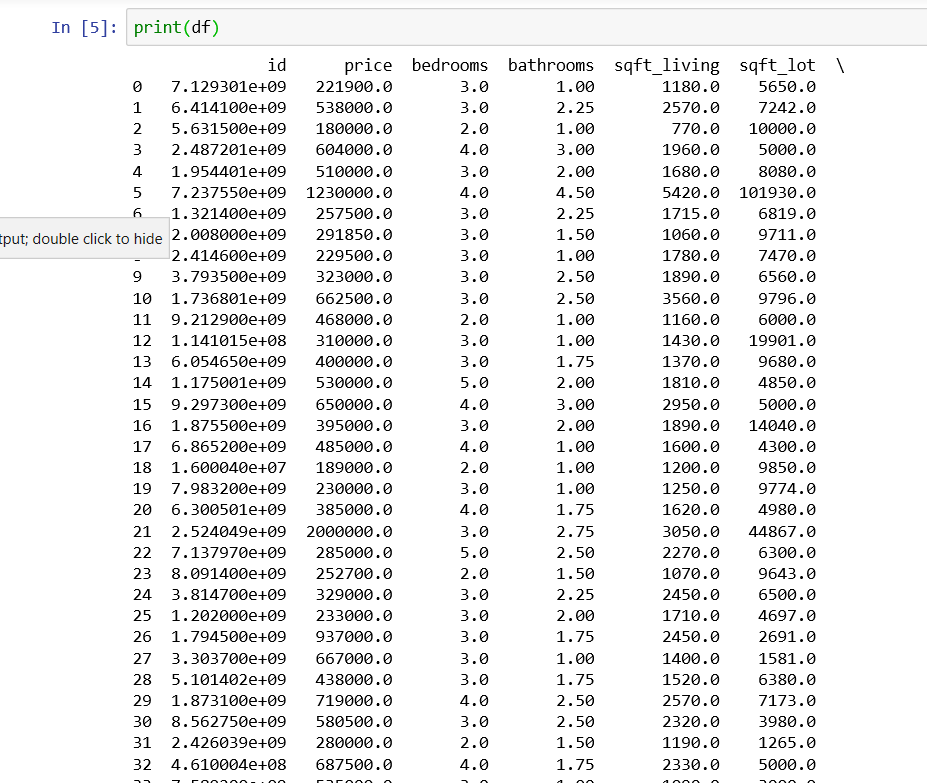
**Matplotlib** is a Python 2D plotting library.

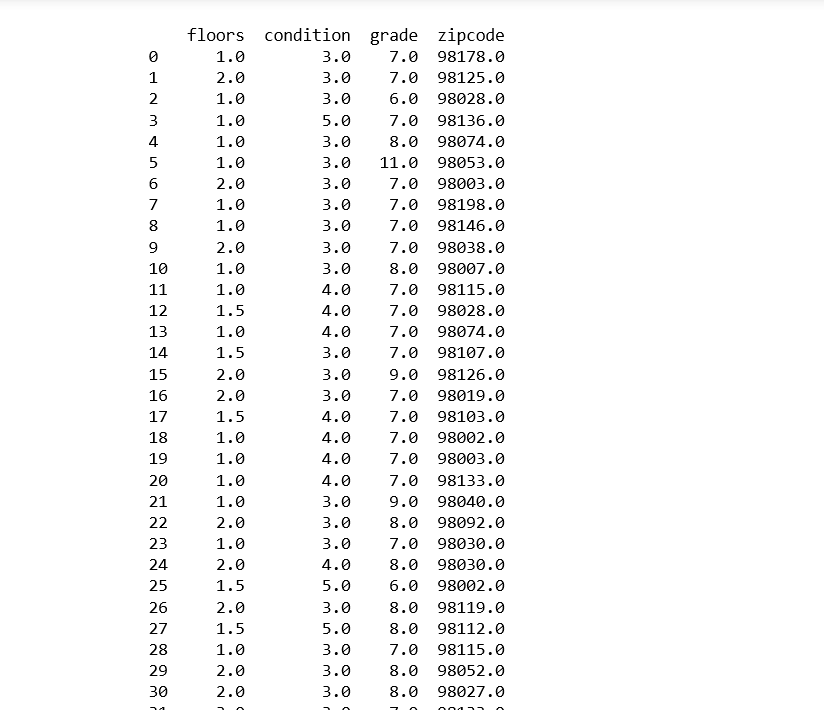
The **pandas** library, easy-to-use data structures and data analysis tools for the python programming language. 

**STEP 2: READ THE DATASET**

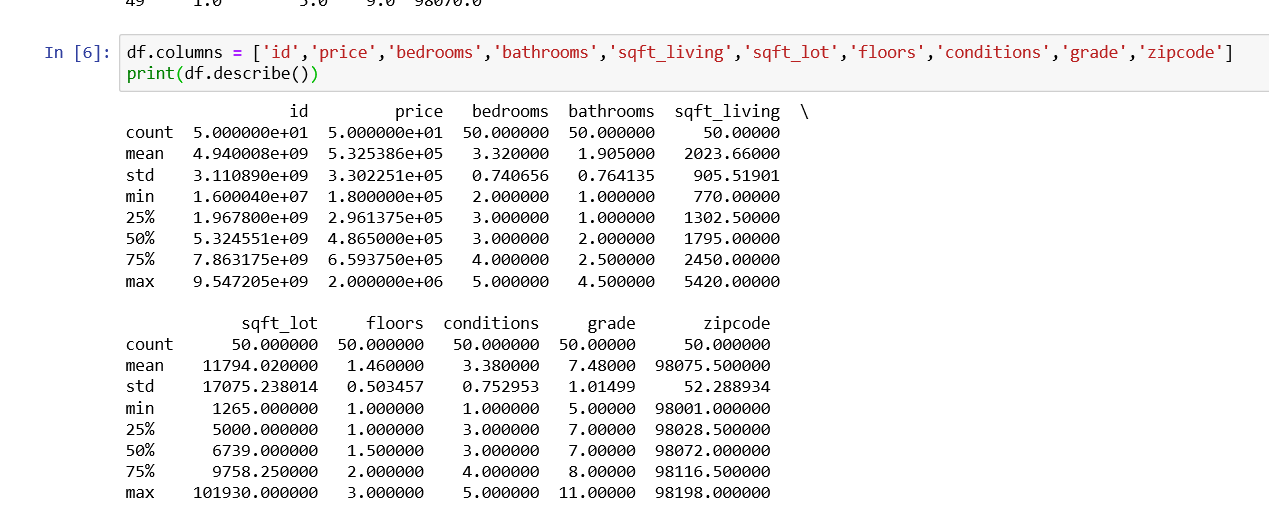








The function **df.describe()** will provide the descriptive statistics of the dataset.

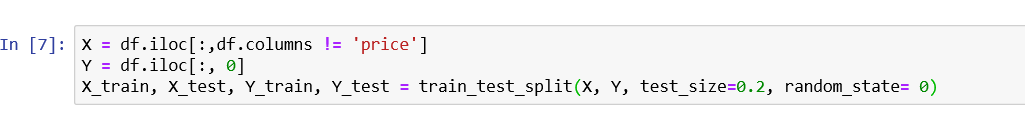


**STEP 3**: **SEPARATE THE INPUT FEATURES AND OUTPUT FROM DATASET**

X = df.iloc[:,df.columns != 'price']

Y = df.iloc[:, 0]

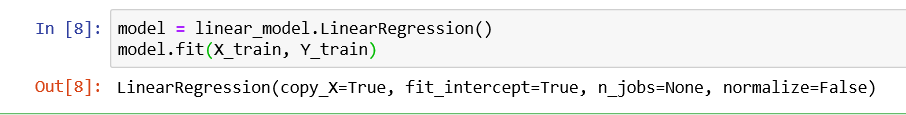
X\_train, X\_test, Y\_train, Y\_test = train\_test\_split(X, Y, test\_size=0.2, random\_state= 0)

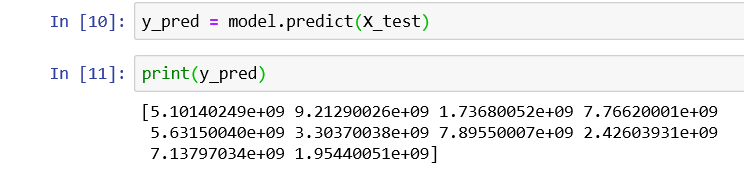
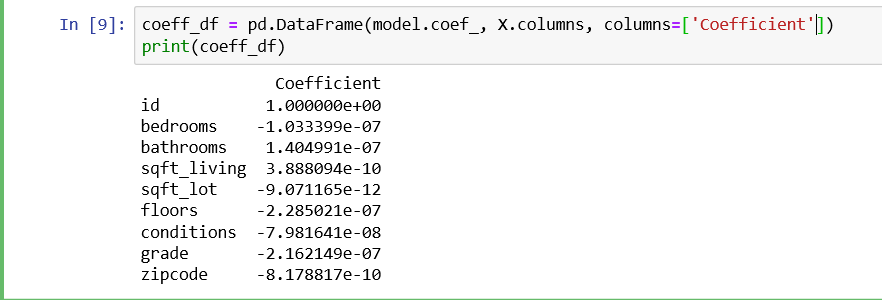


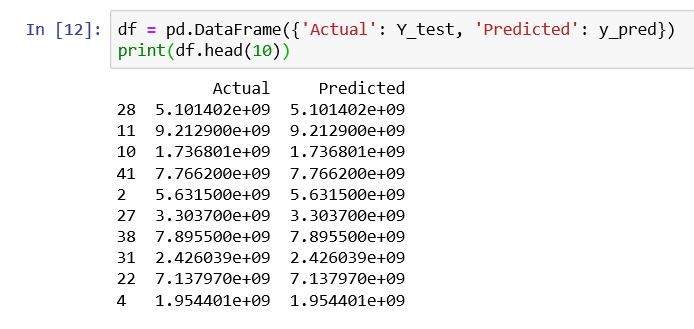
In our dataset price is the output variable, which we need to find out the price of houses. So, we consider all columns except price as input X and 0th column(price) as output Y.

In train\_test\_split() function, we used test\_size as 0.2, which means we use 20% of the data for testing and the remaining 80% for training.

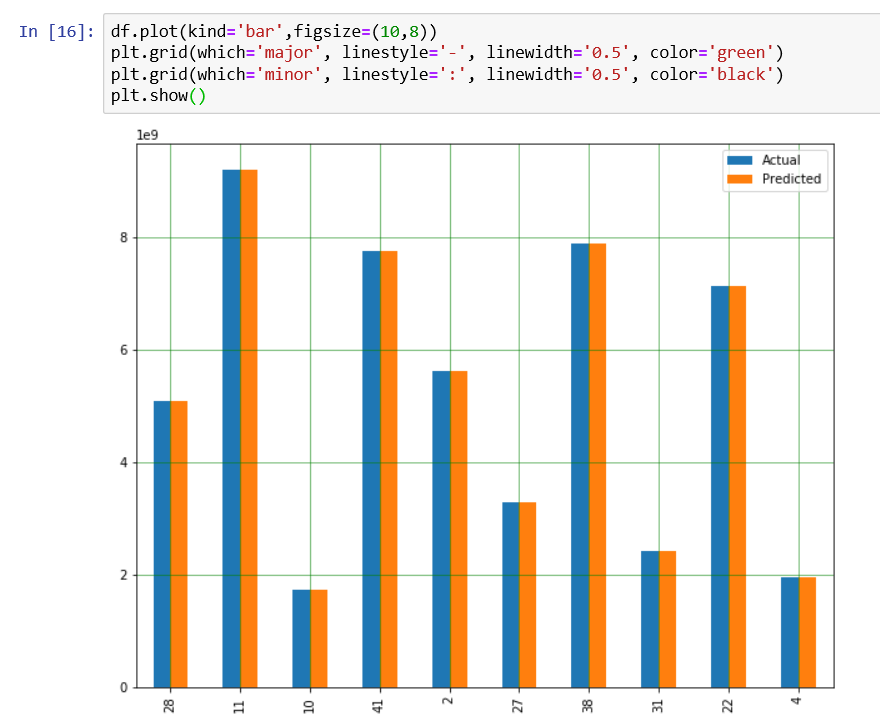
**STEP 4:TRAIN THE MODEL**

we can identify what are the features that have high impact over the output Y. To find that, regression model has to find the most optimal coefficients for all the attributes.

**STEP 5: TO PREDICT THE OUTUT**To check the difference between the predicted value and actual value



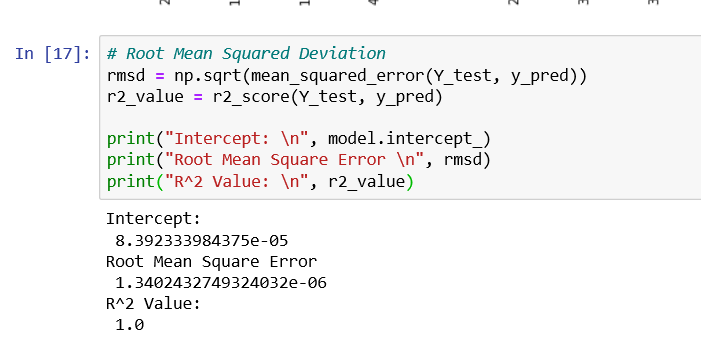
Let’s plot the actual vs predicted value graph



As it is evident from the graph that predicted value and actual value is almost same.

**STEP 6: EVALUATION**

Evaluation can explain how best fit the model by analyzing the R² value and RMSE(Root Mean Square Error).



Here the R² value is 1, which shows the model is almost accurate and can make good predictions. R² value can range from 0 to 1. As the R² value close to 1, the model will make better predictions.