**House Price Prediction Model**

**Group Members –**

**1) Ajay Bhadoria (RA1811026010064)**

**2) Abhimanyu Kejriwal (RA1811026010076)**

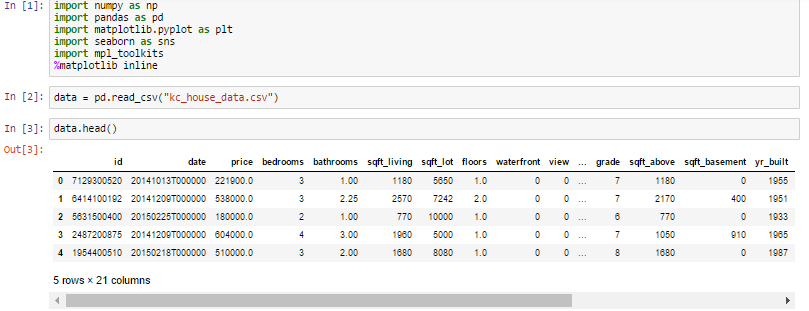
**CSE- AI&ML**

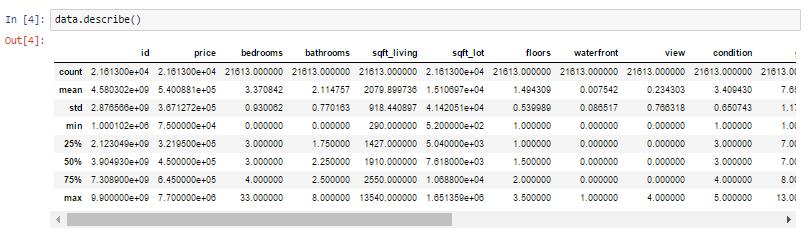
**SECTION – I1**

**BATCH – 1**

**AIM –** Amid the pandemic, it has become difficult for people to step out of their comfort zone and go to places for buying a house. So keeping that in mind, we have developed a model where a person can compare the price of different houses based on different criteria in a particular location.

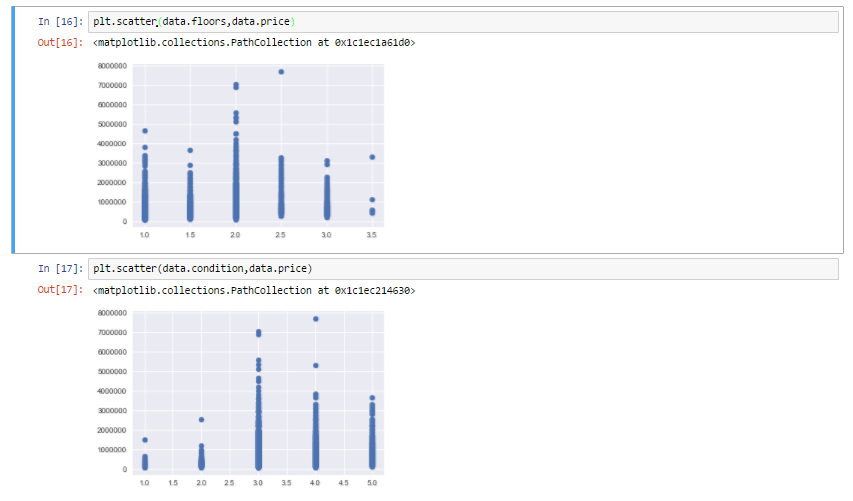
First thing first , we import our libraries and dataset and then we see the head of the data to know how the data looks like and use describe function to see the percentile’s and other key statistics.





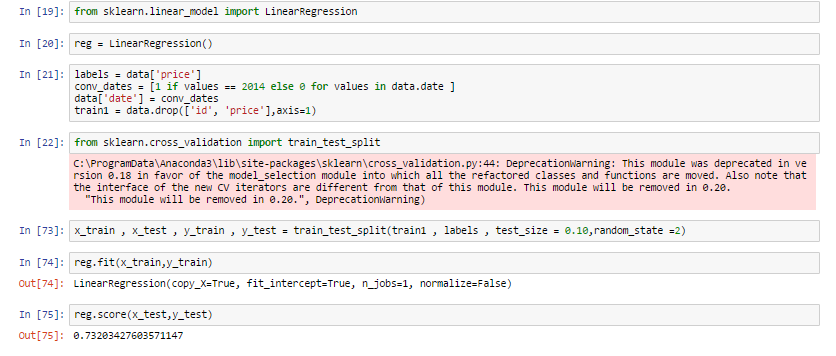
**Factors affecting the price –**

1. House bedrooms and count
2. How houses are placed ?
3. Price vs Square feet and Price vs Longitude
4. Total sqft including basement vs price and waterfront vs price
5. Floors vs Price and condition vs Price



As we can see from all the above representation that many factors are affecting the prices of the house , like square feet which increases the price of the house and even location influencing the prices of the house.Now that we are familiar with all these representation and can tell our own story let us move and create a model to which would predict the price of the house based upon the other factors such as square feet , water front etc . We are going to see what is linear regression and how do we do it ?

**CODE -**

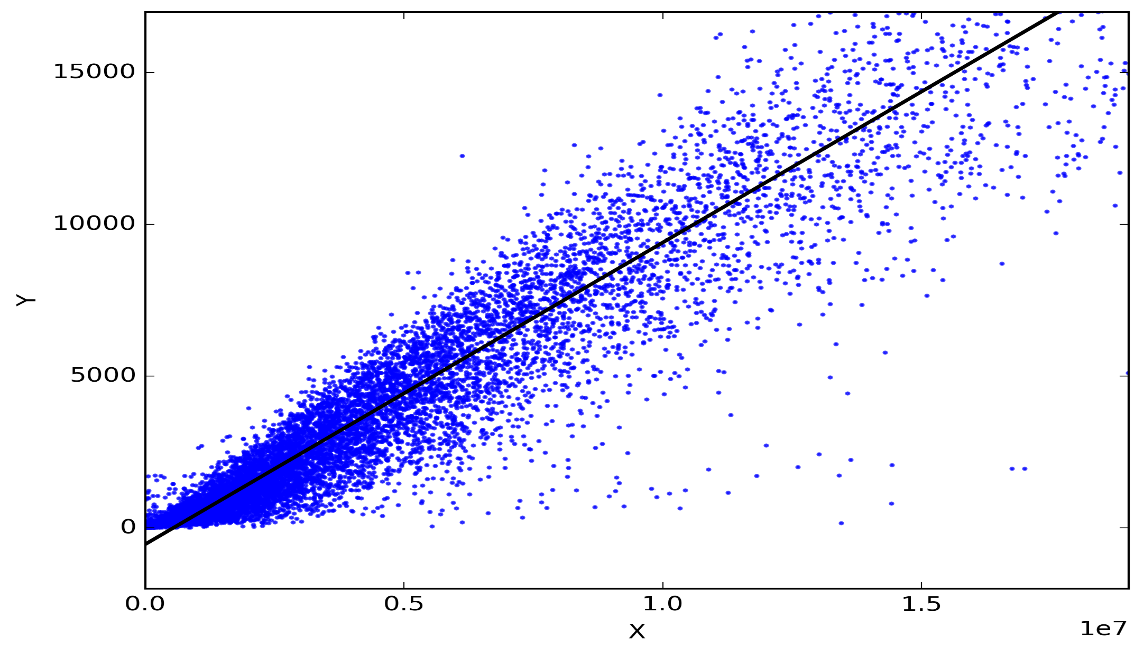


We use train data and test data , train data to train our machine and test data to see if it has learnt the data well or not. Before anything , I want everyone to remember that the machine is the student and train data is the syllabus and test data is the exam. we see how much the machine has scored and if it scores well are model is successful.

1. We import our dependencies , for linear regression we use sklearn (built in python library) and import linear regression from it.
2. We then initialize Linear Regression to a variable reg.
3. Now we know that prices are to be predicted , hence we set labels (output) as price columns and we also convert dates to 1’s and 0’s so that it doesn’t influence our data much . We use 0 for houses which are new that is built after 2014.
4. We again import another dependency to split our data into train and test.
5. I’ve made my train data as 90% and 10% of the data to be my test data , and randomized the splitting of data by using random\_state.
6. So now , we have train data , test data and labels for both let us fit our train and test data into linear regression model.
7. After fitting our data to the model we can check the score of our data ie , prediction. in this case the prediction is 73%

The accuracy of the model is lower than our aim of 85. So how do we achieve that 85% target ? We use a different method , which is very important for weak prediction models such as this. This might seem to be a bit advanced but if understood is a really brilliant tool to enable better predictions. For building a prediction model , many experts use gradient boosting regression , so what is gradient boosting ? It is a machine learning technique for regression and classification problems, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees.

**OUTPUT -**



Regression works on the line equation , y=mx+c , trend line is set through the data points to predict the outcome. The variable we are predicting is called the criterion variable and is referred to as Y. The variable we are basing our predictions on is called the predictor variable and is referred to as X. When there is only one predictor variable, the prediction method is called **Simple Regression. and if multiple predictor variable are present then multiple regression.**