

Housing Price Problem

- Q1) import the libraries (pandas,matplotlib,seaborn,scikit,numpy)
- Q2) read the csv file and store it in a dataframe
- Q3) check out the top 5 data in your dataframe using the head() function of pandas
- Q4) check out the bottom 5 data in your dataframe using tail() function
- Q5) check out the info of your dataframe using the info function of pandas
- Q6) check out the summary of dataframe using describe method of pandas
- Q7) Use the pairplot method of seaborn and check out the graph
- Q8) Plot the graph for Price vs Area
- Q9) Plot the graph for House Age vs Price
- Q10) Plot the graph for population vs income.
- Q11) Plot the graph for Area vs Rooms
- Q12) Use distplot on price column and observe the distribution
- Q13) Use distplot on house age column and observe the distribution
- Q14) Use distplot on population column and observe the distribution
- Q15) Observe the pic below and code accordingly

Training a Linear Regression Model

Let's now begin to train our regression model! We will need to first split up our data into an X array that contains the features to train on, and a y array with the target variable, in this case the Price column. We will toss out the Address column because it only has text info that the linear regression model can't use.

X and y arrays

```
In [264]: 1 X = USAhousing[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',  
2               'Avg. Area Number of Bedrooms', 'Area Population']]  
3 y = USAhousing['Price']
```

Q16)

Train Test Split

Now let's split the data into a training set and a testing set. We will train our model on the training set and then use the test set to evaluate the model.

```
1 from sklearn.model_selection import train_test_split  
  
1 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=101)
```

Q17)

Creating and Training the Model

```
: 1 from sklearn.linear_model import LinearRegression
: 1 lm = LinearRegression()
: 1 lm.fit(X_train,y_train)
* LinearRegression(copy_X=True fit_intercept=True n_jobs=1 normalize=False)
```

Q18) observe your model when you code and check your output

Model Evaluation

Let's evaluate the model by checking out it's coefficients and how we can interpret them.

```
l: 1 # print the intercept
2 print(lm.intercept_)
-2640159.79685
l: 1 coeff_df = pd.DataFrame(lm.coef_,X.columns,columns=['Coefficient'])
2 coeff_df
```

```
l:

```

	Coefficient
Avg. Area Income	21.528276
Avg. Area House Age	164883.282027
Avg. Area Number of Rooms	122368.678027
Avg. Area Number of Bedrooms	2233.801864
Area Population	15.150420