

```

# MACHINE LEARNING REGRESSION HOUSING DATA
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

dataset = pd.read_csv(r"C:\Users\Gopi Reddy\NIT7PM\mar\26th- mlr\26th- mlr\MLR\House_data.csv")

x = dataset.drop(columns = ['price', 'id', 'date'])

y = dataset.iloc[:, 2]

from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y, test_size=0.2, random_state=0)

from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(x_train, y_train)

y_pred = regressor.predict(x_test)

m_slope = regressor.coef_
print(m_slope)

c_inter = regressor.intercept_
print(c_inter)

x = np.append(arr = np.ones((21613,1)).astype(int), values= x, axis= 1)

import statsmodels.api as sm
x_opt = x[:,[0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17]]
#OrdinaryLeastSquares
regressor_OLS = sm.OLS(endog=y, exog=x_opt).fit()
regressor_OLS.summary()

# Backward elimination based on p-value = 0.05

import statsmodels.api as sm
x_opt = x[:,[0,1,2,3,5,6,7,8,9,10,11,12,13,14,15,16,17]]
#OrdinaryLeastSquares
regressor_OLS = sm.OLS(endog=y, exog=x_opt).fit()
regressor_OLS.summary()

#understanding the distribution with seaborn

with sns.plotting_context("notebook",font_scale=2.5):
    g = sns.pairplot(dataset[['sqft_lot','sqft_above','price','sqft_living','bedrooms']],
                      hue='bedrooms', palette='tab20',size=6)
g.set(xticklabels=[]);

```