

```

import numpy as np
import matplotlib.pyplot as plt

import pandas as pd
import seaborn as sns

%matplotlib inline

boston_dataset = pd.read_csv("/content/BostonHousing.csv")

boston_dataset.head()

boston_dataset.isnull().sum()

correlation_matrix = boston_dataset.corr().round(2)
# annot = True to print the values inside the square
sns.heatmap(data=correlation_matrix, annot=True)

plt.figure(figsize=(20, 5))

features = ['lstat', 'rm']
target = boston_dataset['medv']

for i, col in enumerate(features):
    plt.subplot(1, len(features) , i+1)
    x = boston_dataset[col]
    y = target
    plt.scatter(x, y, marker='o')
    plt.title(col)
    plt.xlabel(col)
    plt.ylabel('medv')

X = pd.DataFrame(np.c_[boston_dataset['lstat'], boston_dataset['rm']],
columns = ['lstat', 'rm'])
Y = boston_dataset['medv']

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=5)
print(X_train.shape)
print(X_test.shape)
print(Y_train.shape)
print(Y_test.shape)

from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import statsmodels.api as sm

lin_model = LinearRegression()
lin_model.fit(X_train, Y_train)

```

```
model = sm.OLS(Y_train, X_train).fit()  
print(model.summary())
```