4	**4**	**4**	**4**	**4**	**20**	

Group A
Assignment No: 4

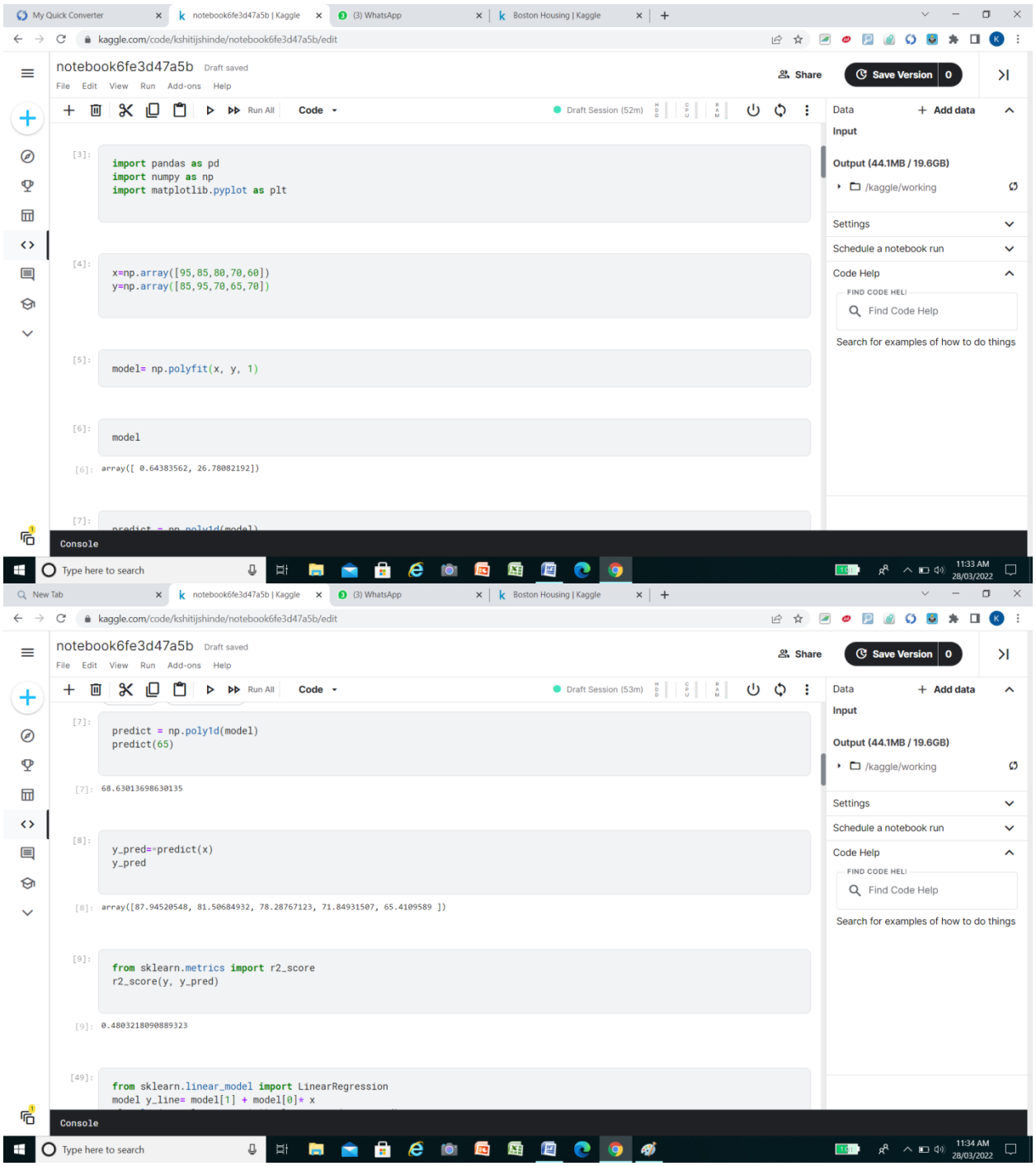
Title of the Assignment:

Create a Linear Regression Model using Python/R to predict home prices using Boston Housing Dataset (https://www.kaggle.com/c/boston-housing).

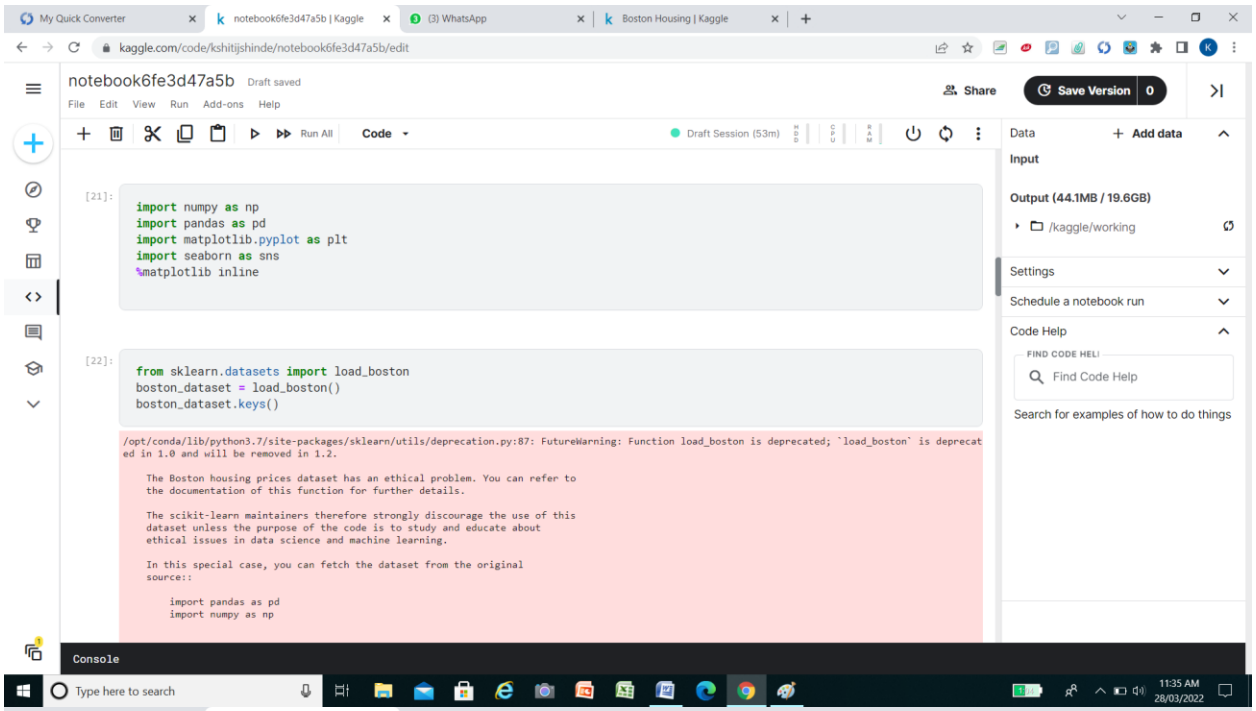
The Boston Housing dataset contains information about various houses in Boston through different parameters. There are 506 samples and 14 feature variables in this dataset.

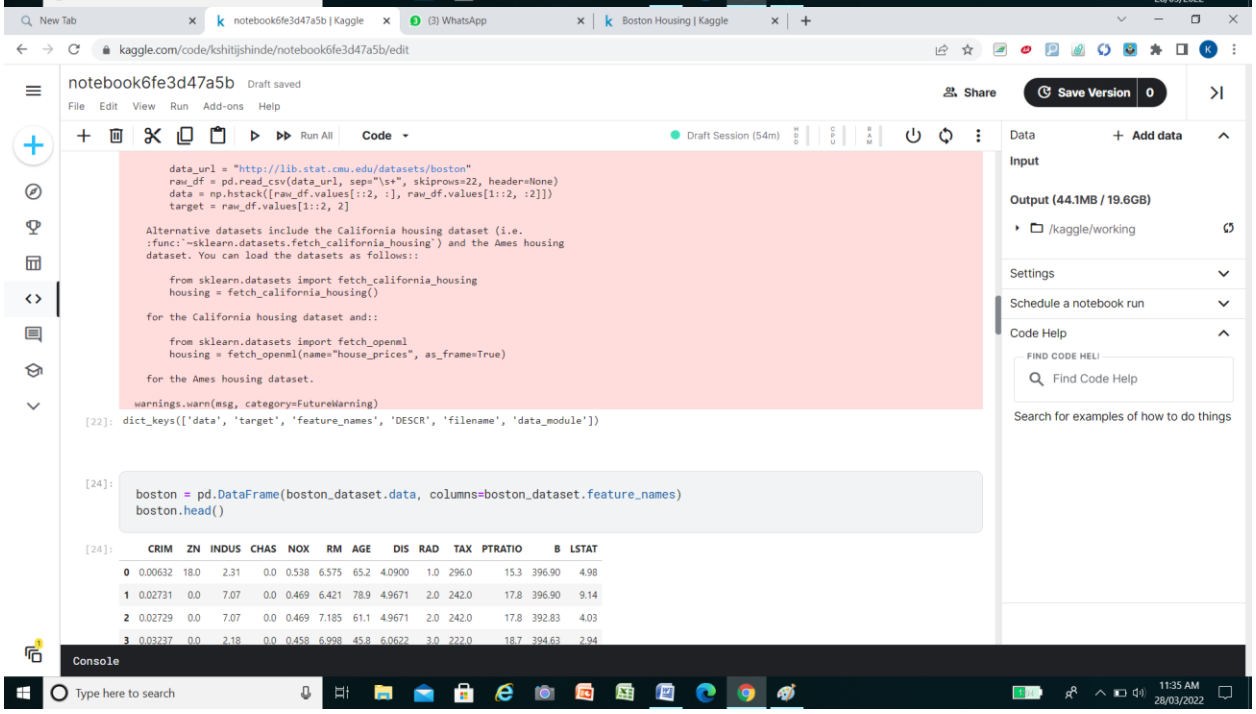
The objective is to predict the value of prices of the house using the given features.

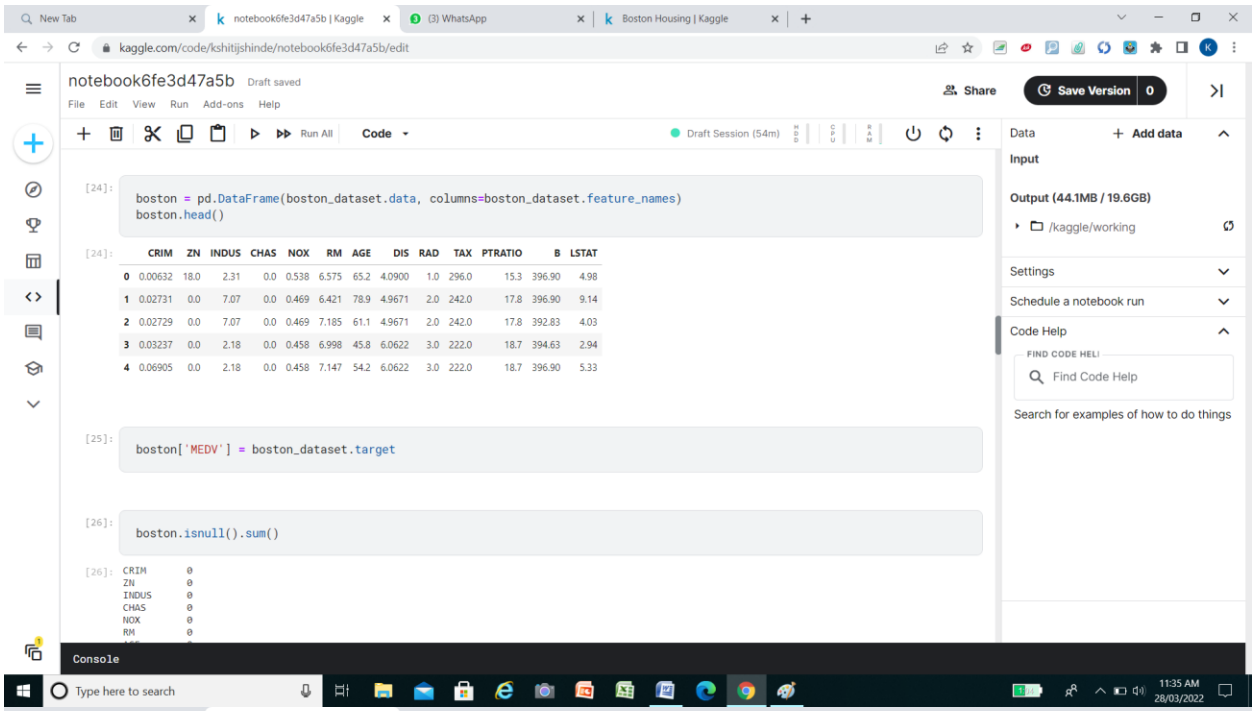
OUTPUT :

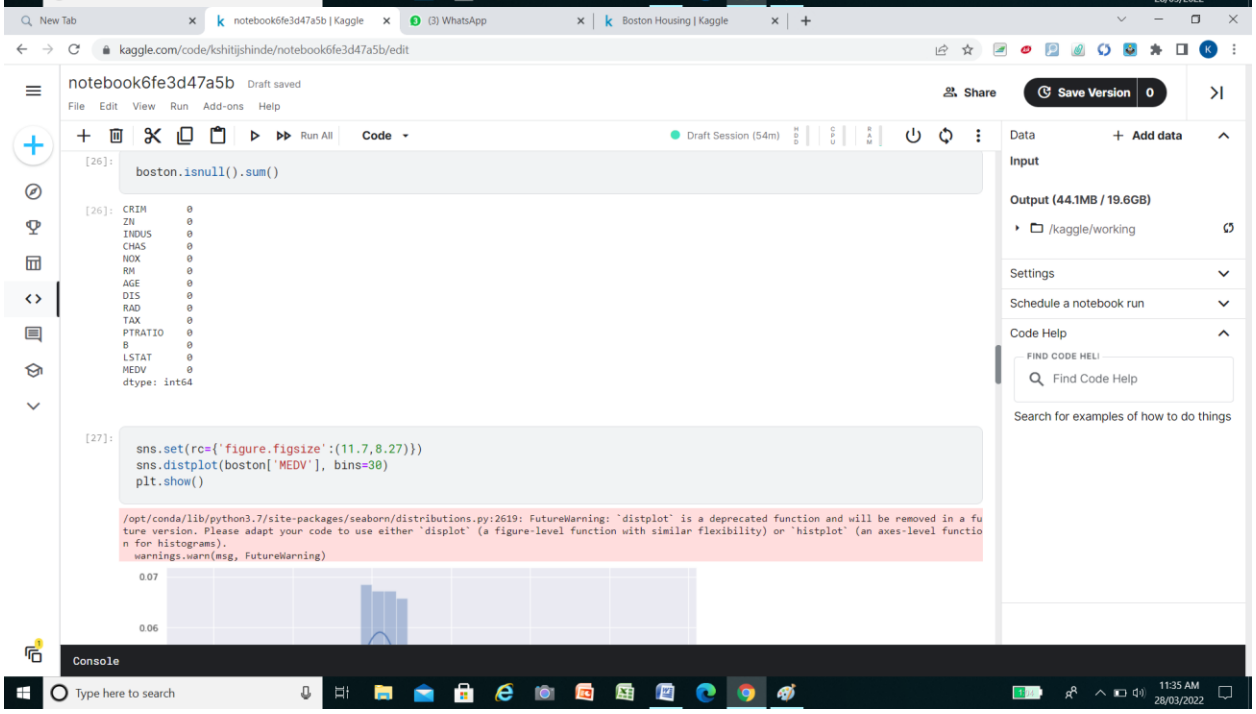

My Quick Converter
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (52m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[3]:
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
[4]:
x=np.array([95,85,80,70,60])
y=np.array([85,95,70,65,70])
[5]:
model= np.polyfit(x, y, 1)
[6]:
model
[6]: array([0.64383562, 26.78082192])
[7]:
Console
Type here to search
11:33 AM
28/03/2022

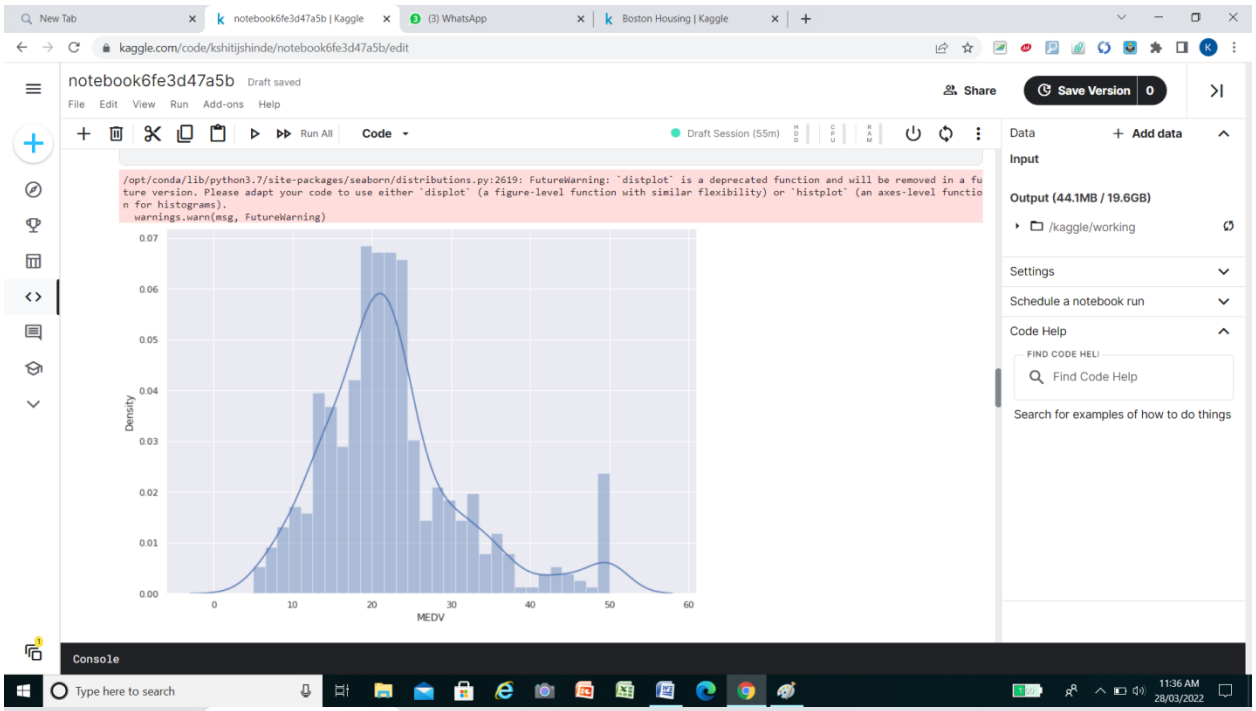
New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (53m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[7]:
predict = np.poly1d(model)
predict(65)
[7]: 68.63013698630135
[8]:
y_pred=predict(x)
y_pred
[8]: array([87.94520548, 81.50684932, 78.28767123, 71.84931507, 65.4109589])
[9]:
from sklearn.metrics import r2_score
r2_score(y, y_pred)
[9]: 0.4803218090889323
[49]:
from sklearn.linear_model import LinearRegression
model y_line= model[1] + model[0]* x
Console
Type here to search
11:34 AM
28/03/2022

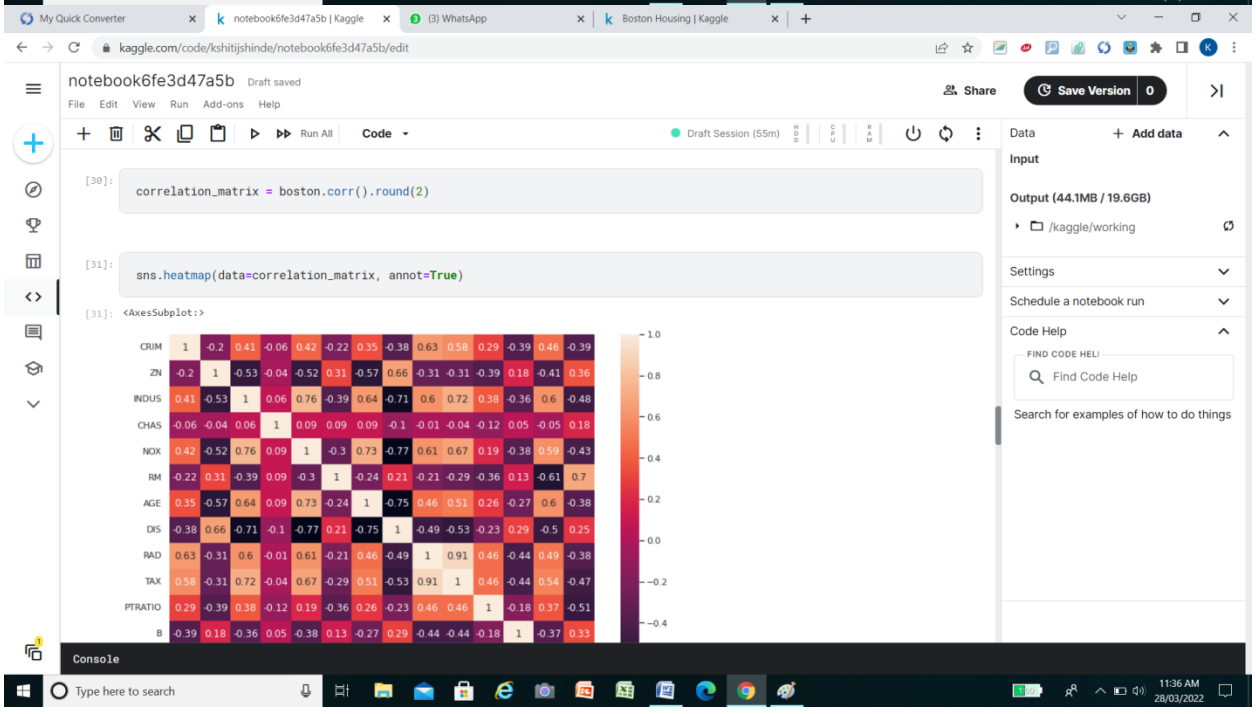

My Quick Converter
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (53m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[21]:
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
[22]:
from sklearn.datasets import load_boston
boston_dataset = load_boston()
boston_dataset.keys()
/opt/conda/lib/python3.7/site-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function load_boston is deprecated; `load_boston` is deprecated in 1.0 and will be removed in 1.2.
The Boston housing prices dataset has an ethical problem. You can refer to
the documentation of this function for further details.
The scikit-learn maintainers therefore strongly discourage the use of this
dataset unless the purpose of the code is to study and educate about
ethical issues in data science and machine learning.
In this special case, you can fetch the dataset from the original
source::
import pandas as pd
import numpy as np
Console
Type here to search
11:35 AM
28/03/2022


New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (54m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
data_url = "http://lib.stat.cmu.edu/datasets/boston"
raw_df = pd.read_csv(data_url, sep="\s+", skiprows=22, header=None)
data = np.hstack([raw_df.values[::2, :], raw_df.values[1::2, :2]])
target = raw_df.values[1::2, 2]
Alternative datasets include the California housing dataset (i.e.
:func:`~sklearn.datasets.fetch_california_housing`) and the Ames housing
dataset. You can load the datasets as follows::
from sklearn.datasets import fetch_california_housing
housing = fetch_california_housing()
for the California housing dataset and::
from sklearn.datasets import fetch_openml
housing = fetch_openml(name="house_prices", as_frame=True)
for the Ames housing dataset.
warnings.warn(msg, category=FutureWarning)
[22]:
dict_keys(['data', 'target', 'feature_names', 'DESCR', 'filename', 'data_module'])
[24]:
boston = pd.DataFrame(boston_dataset.data, columns=boston_dataset.feature_names)
boston.head()
[24]:
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT
0 0.00632 18.0 2.31 0.0 0.538 6.575 65.2 4.0900 1.0 296.0 15.3 396.90 4.98
1 0.02731 0.0 7.07 0.0 0.469 6.421 78.9 4.9671 2.0 242.0 17.8 396.90 9.14
2 0.02729 0.0 7.07 0.0 0.469 7.185 61.1 4.9671 2.0 242.0 17.8 392.83 4.03
3 0.03237 0.0 2.18 0.0 0.458 6.998 45.8 6.0622 3.0 222.0 18.7 394.63 2.94
Console
Type here to search
11:35 AM
28/03/2022


New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (54m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[24]:
boston = pd.DataFrame(boston_dataset.data, columns=boston_dataset.feature_names)
boston.head()
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B LSTAT
0 0.00632 18.0 2.31 0.0 0.538 6.575 65.2 4.0900 1.0 296.0 15.3 396.90 4.98
1 0.02731 0.0 7.07 0.0 0.469 6.421 78.9 4.9671 2.0 242.0 17.8 396.90 9.14
2 0.02729 0.0 7.07 0.0 0.469 7.185 61.1 4.9671 2.0 242.0 17.8 392.83 4.03
3 0.03237 0.0 2.18 0.0 0.458 6.998 45.8 6.0622 3.0 222.0 18.7 394.63 2.94
4 0.06905 0.0 2.18 0.0 0.458 7.147 54.2 6.0622 3.0 222.0 18.7 396.90 5.33
[25]:
boston['MEDV'] = boston_dataset.target
[26]:
boston.isnull().sum()
[26]:
CRIM 0
ZN 0
INDUS 0
CHAS 0
NOX 0
RM 0
Console
Type here to search
11:35 AM
28/03/2022


New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (54m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[26]:
boston.isnull().sum()
[26]:
CRIM 0
ZN 0
INDUS 0
CHAS 0
NOX 0
RM 0
AGE 0
DIS 0
RAD 0
TAX 0
PTRATIO 0
B 0
LSTAT 0
MEDV 0
dtype: int64
[27]:
sns.set(rc={'figure.figsize':(11.7,8.27)})
sns.distplot(boston['MEDV'], bins=30)
plt.show()
/opt/conda/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a fu
ture version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level functio
n for histograms).
warnings.warn(msg, FutureWarning)
0.07
0.06
Console
Type here to search
11:35 AM
28/03/2022


notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (55m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
/opt/conda/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
warnings.warn(msg, FutureWarning)
Density
MEDV
Console


notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (55m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[30]:
correlation_matrix = boston.corr().round(2)
[31]:
sns.heatmap(data=correlation_matrix, annot=True)
[31]: <AxesSubplot:>
CRIM ZN INDUS CHAS NOX RM AGE DIS RAD TAX PTRATIO B
Console

My Quick Converter | notebook6fe3d47a5b | Kaggle | (3) WhatsApp | Boston Housing | Kaggle

kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit

notebook6fe3d47a5b Draft saved

File Edit View Run Add-ons Help

Share | Save Version 0

Run All | Code | Draft Session (55m)

Data | Add data

Input

Output (44.1MB / 19.6GB)

/kaggle/working

Settings

Schedule a notebook run

Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

[31]:
```
sns.heatmap(data=correlation_matrix, annot=True)
```

[31]: <AxesSubplot:>

Console

Type here to search

11:37 AM 28/03/2022

New Tab | notebook6fe3d47a5b | Kaggle | (3) WhatsApp | Boston Housing | Kaggle

kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit

notebook6fe3d47a5b Draft saved

File Edit View Run Add-ons Help

Share | Save Version 0

Run All | Code | Draft Session (56m)

Data | Add data

Input

Output (44.1MB / 19.6GB)

/kaggle/working

Settings

Schedule a notebook run

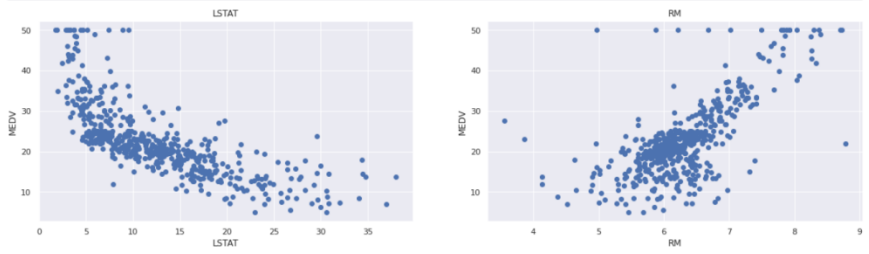
Code Help

FIND CODE HELP

Find Code Help

Search for examples of how to do things

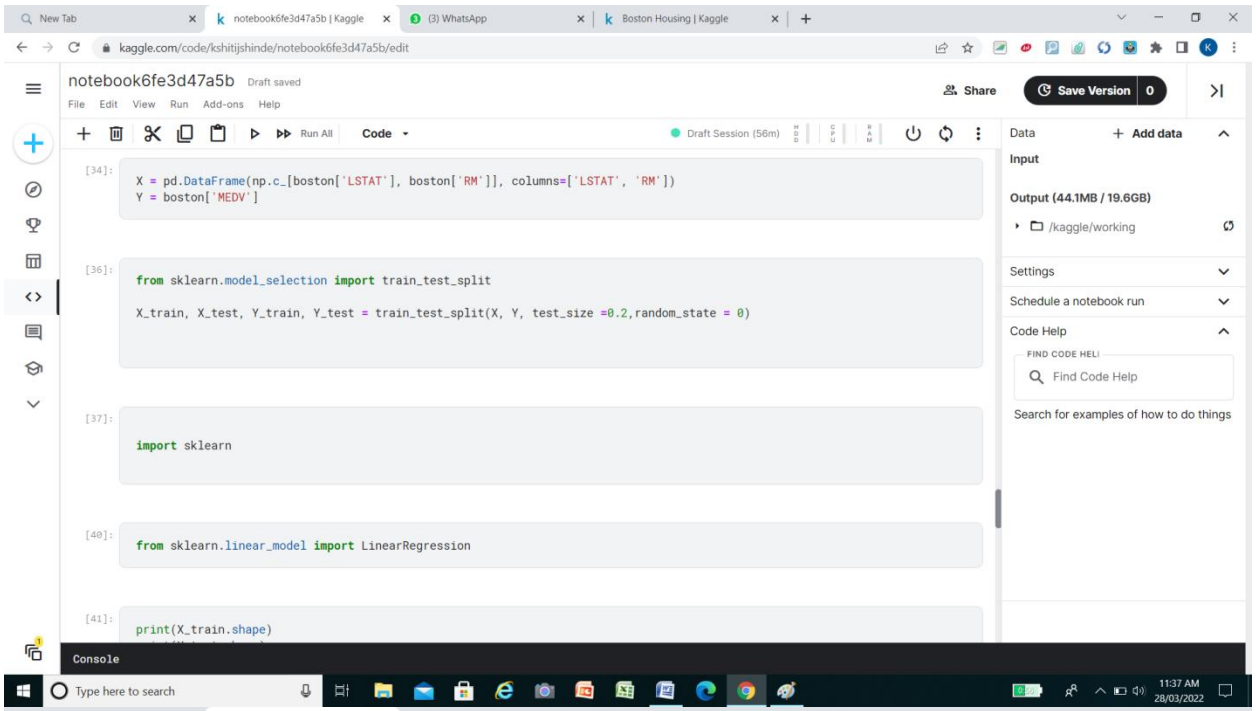
[33]:
```
plt.figure(figsize=(20, 5))
features = ['LSTAT' , 'RM']
target = boston['MEDV']
for i, col in enumerate(features):
    plt.subplot(1, len(features), i+1)
    x = boston[col]
    y = target
    plt.scatter(x, y, marker='o')
    plt.title(col)
    plt.xlabel(col)
    plt.ylabel('MEDV')
```

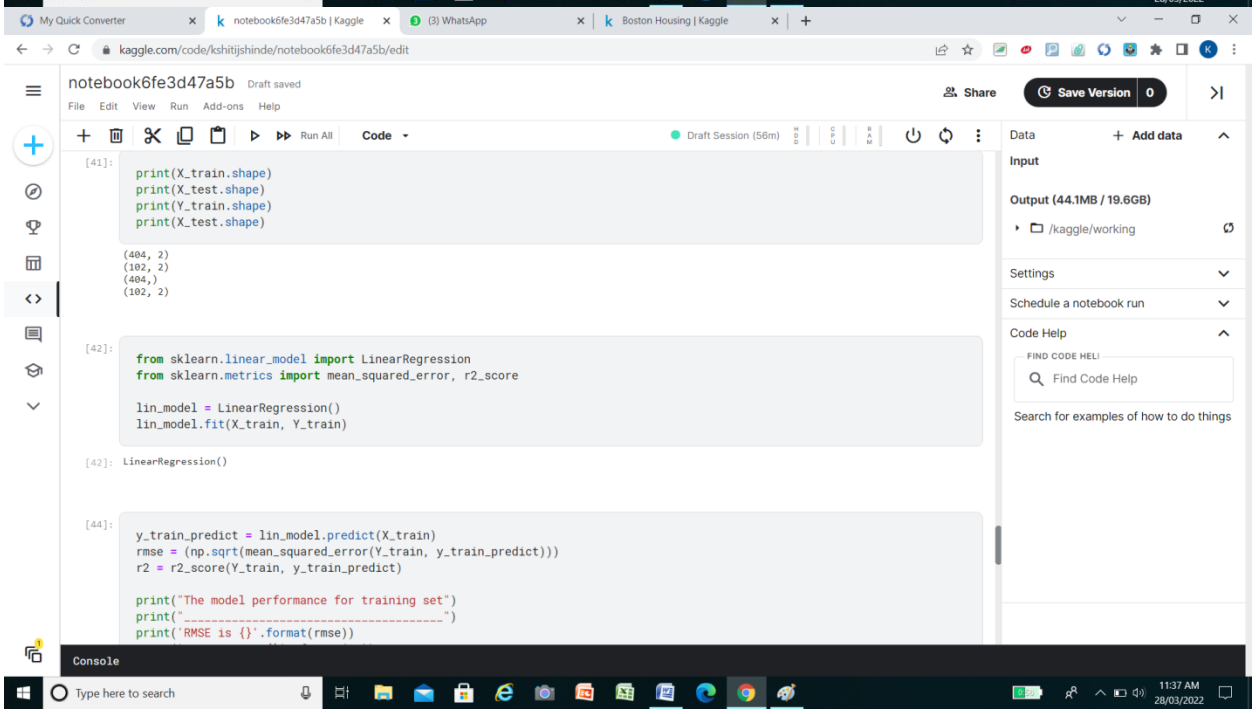


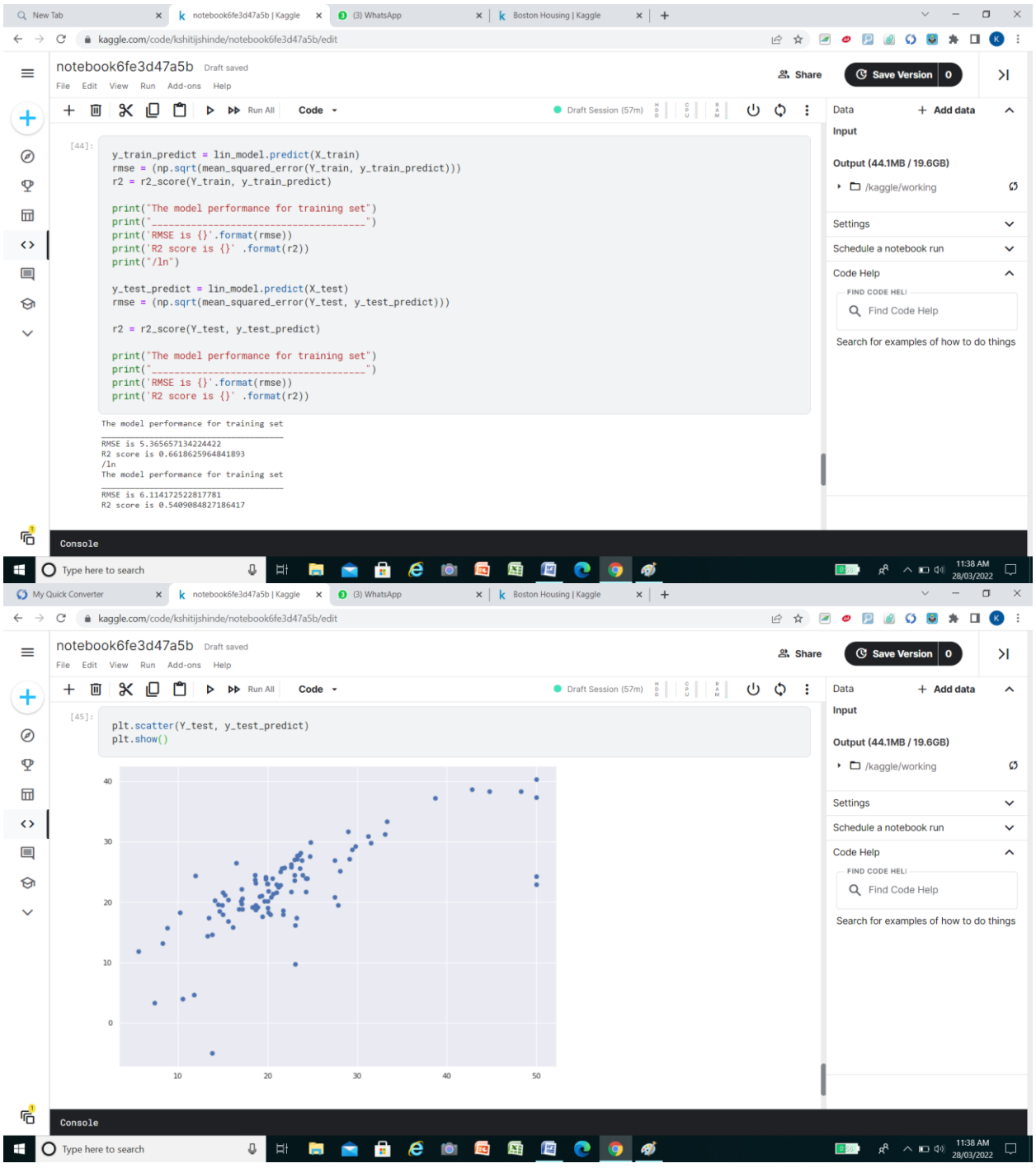
Console

Type here to search

11:37 AM 28/03/2022


New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (56m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[34]:
X = pd.DataFrame(np.c_[boston['LSTAT'], boston['RM']], columns=['LSTAT', 'RM'])
Y = boston['MEDV']
[36]:
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)
[37]:
import sklearn
[40]:
from sklearn.linear_model import LinearRegression
[41]:
print(X_train.shape)
Console
Type here to search
11:37 AM
28/03/2022


My Quick Converter
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (56m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[41]:
print(X_train.shape)
print(X_test.shape)
print(Y_train.shape)
print(X_test.shape)
(404, 2)
(102, 2)
(404,)
(102, 2)
[42]:
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
lin_model = LinearRegression()
lin_model.fit(X_train, Y_train)
[42]: LinearRegression()
[44]:
y_train_predict = lin_model.predict(X_train)
rmse = (np.sqrt(mean_squared_error(Y_train, y_train_predict)))
r2 = r2_score(Y_train, y_train_predict)
print("The model performance for training set")
print("--------------------------------------")
print('RMSE is {}'.format(rmse))
Console
Type here to search
11:37 AM
28/03/2022


New Tab
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (57m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[44]:
y_train_predict = lin_model.predict(X_train)
rmse = (np.sqrt(mean_squared_error(Y_train, y_train_predict)))
r2 = r2_score(Y_train, y_train_predict)
print("The model performance for training set")
print("--------------------------------------")
print('RMSE is {}'.format(rmse))
print('R2 score is {}' .format(r2))
print("/ln")
y_test_predict = lin_model.predict(X_test)
rmse = (np.sqrt(mean_squared_error(Y_test, y_test_predict)))
r2 = r2_score(Y_test, y_test_predict)
print("The model performance for training set")
print("--------------------------------------")
print('RMSE is {}'.format(rmse))
print('R2 score is {}' .format(r2))
The model performance for training set

RMSE is 5.365657134224422
R2 score is 0.6618625964841893
/ln
The model performance for training set

RMSE is 6.114172522817781
R2 score is 0.5409084827186417
Console
Type here to search
11:38 AM
28/03/2022
My Quick Converter
notebook6fe3d47a5b | Kaggle
(3) WhatsApp
Boston Housing | Kaggle
kaggle.com/code/kshitijshinde/notebook6fe3d47a5b/edit
notebook6fe3d47a5b
Draft saved
File Edit View Run Add-ons Help
Share
Save Version 0
Run All
Code
Draft Session (57m)
Data
Add data
Input
Output (44.1MB / 19.6GB)
/kaggle/working
Settings
Schedule a notebook run
Code Help
FIND CODE HELP
Find Code Help
Search for examples of how to do things
[45]:
plt.scatter(Y_test, y_test_predict)
plt.show()
40
30
20
10
0
10
20
30
40
50
Console
Type here to search
11:38 AM
28/03/2022