

# IBM ASSIGNMENT 4

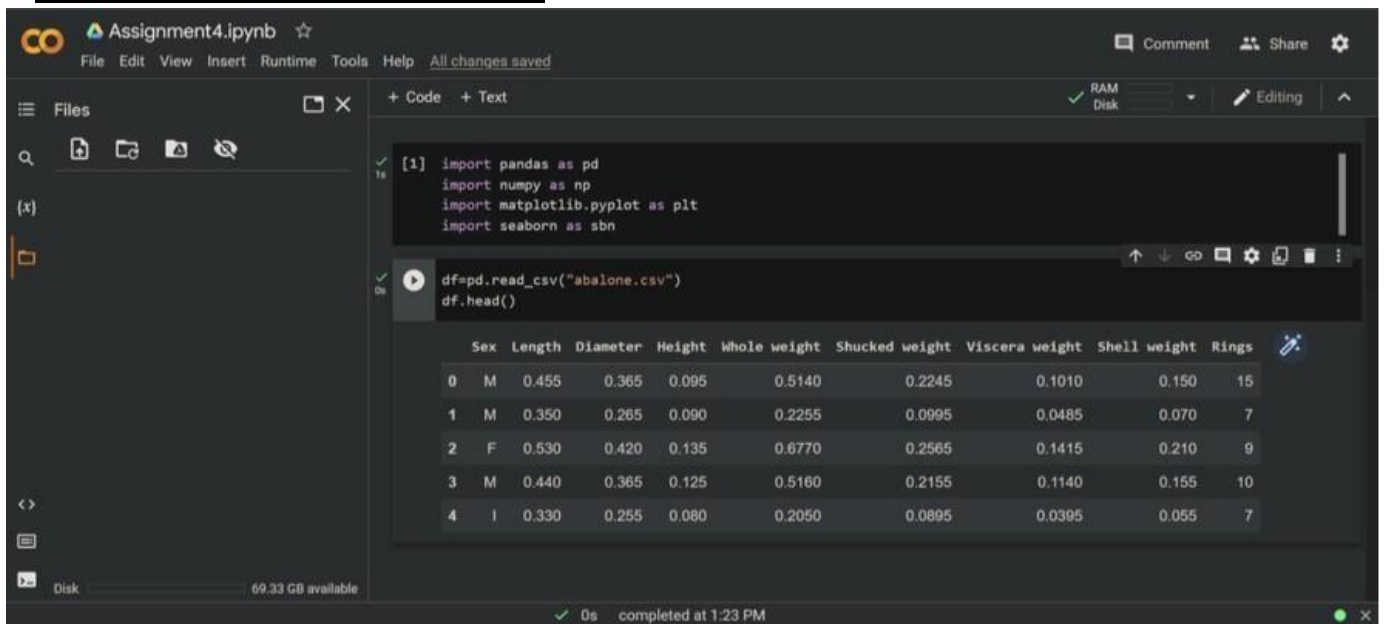
Name: R. Raja

Register No: 961819104067

## CHALLENGE:

Abalone Age Prediction

## LOADING THE DATASET:



The screenshot shows a Jupyter Notebook titled "Assignment4.ipynb". The code cell contains the following Python code:

```
[1] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv("abalone.csv")
df.head()
```

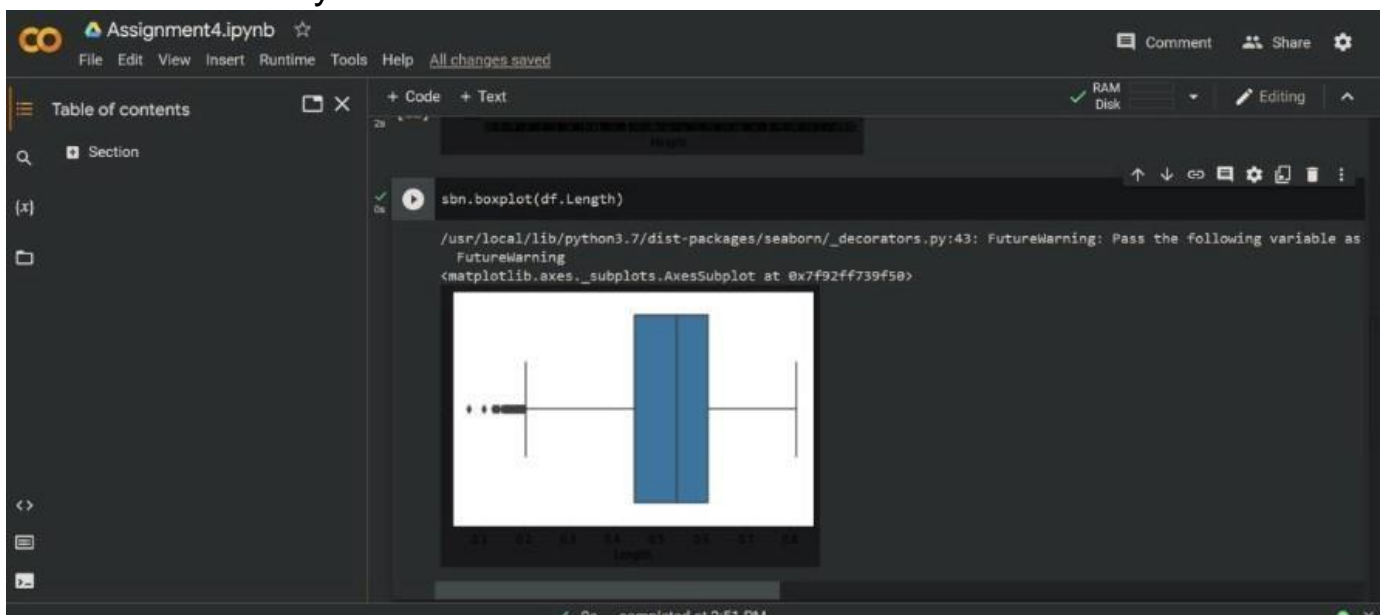
The output of the code is a preview of the first five rows of the "abalone.csv" dataset:

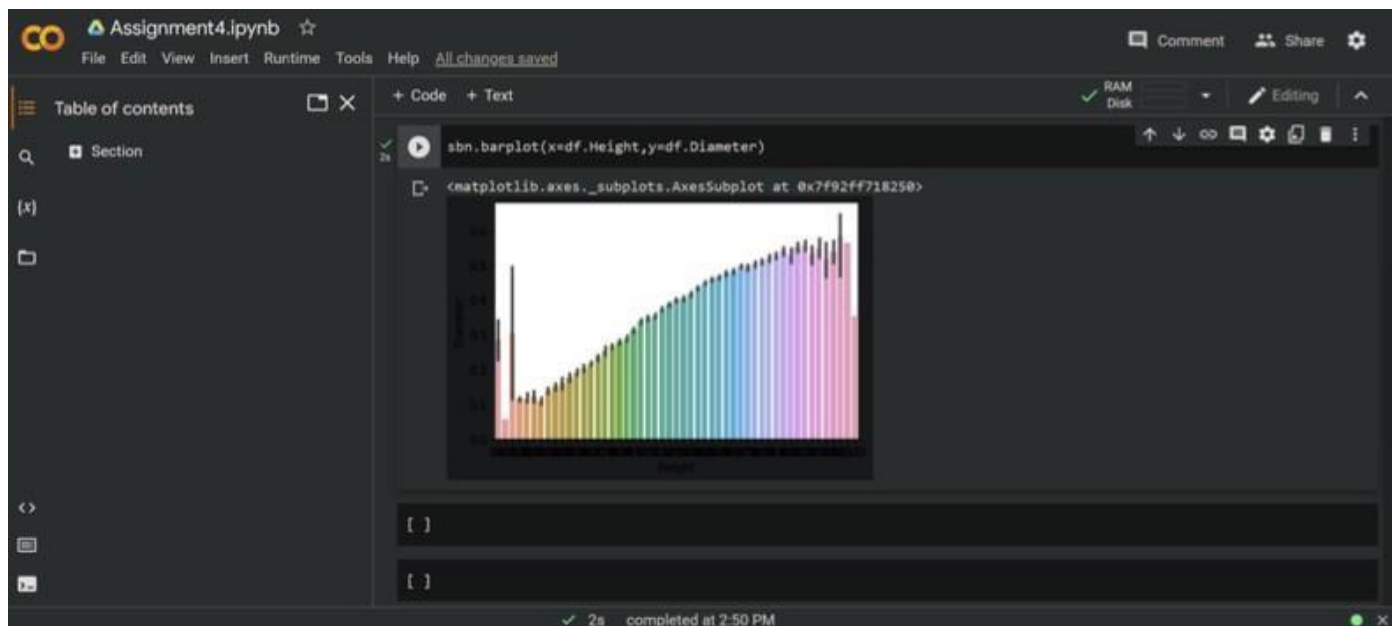
	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	M	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

## VISUALIZATIONS:

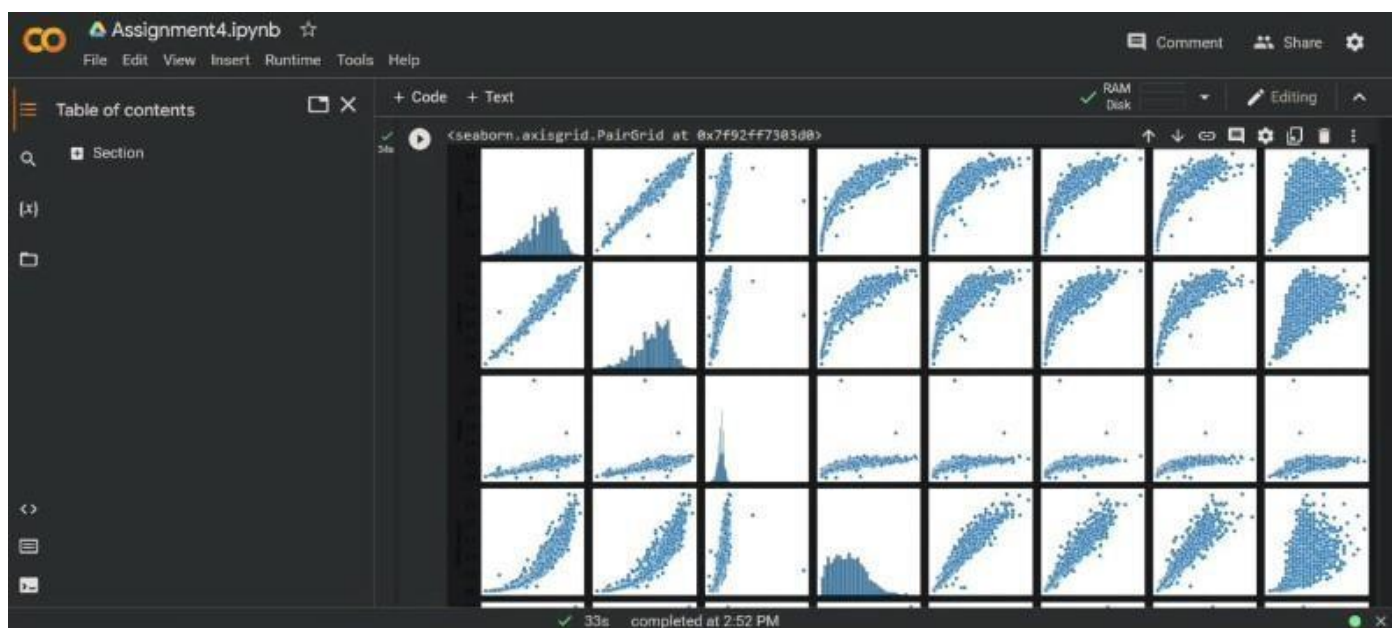
Univariate Analysis

Bi-Variate Analysis





## Multi-Variate Analysis



Perform descriptive Analysis on datasets

This screenshot shows a Jupyter Notebook titled 'Assignment4.ipynb'. The left sidebar contains a 'Table of contents' and a search bar. The main area displays three code cells with their outputs:

```
[15] df['Length'].mode()
0    0.550
1    0.625
dtype: float64
```

```
[17] df['Height'].mean()
0.13951639932966242
```

```
[20] df.count()
Sex            4177
Length         4177
Diameter       4177
Height         4177
Whole weight   4177
Shucked weight 4177
Viscera weight 4177
Shell weight   4177
Rings          4177
dtype: int64
```

The status bar at the bottom indicates '0s completed at 2:56 PM'.

This screenshot shows the same Jupyter Notebook at a later stage. The code cells and their outputs are:

```
Rings            4177
dtype: int64
```

```
[23] df['Shell weight'].sum()
997.5964999999999
```

```
[24] df['Rings'].product()
0
```

```
[25] df['Whole weight'].max()
2.8255
```

Below the code cells, there is a progress bar and a status bar at the bottom indicating '0s completed at 2:59 PM'.

Checking for missing values and deal with them , Finding the outliers and replace them outliers

Assignment4.ipynb

File Edit View Insert Runtime Tools Help Saving...

Table of contents

Section

(x)

[27] df.isna().any()

```
Sex          False
Length       False
Diameter     False
Height       False
Whole weight False
Shucked weight False
Viscera weight False
Shell weight False
Rings        False
dtype: bool
```

3.0

qu1=df.Rings.quantile(0.25)  
qu3=df.Rings.quantile(0.75)  
qr=qu3-qu1  
print(qr)

3.0

0s completed at 3:01 PM

Check for categorical columns and perform encoding

Assignment4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Table of contents

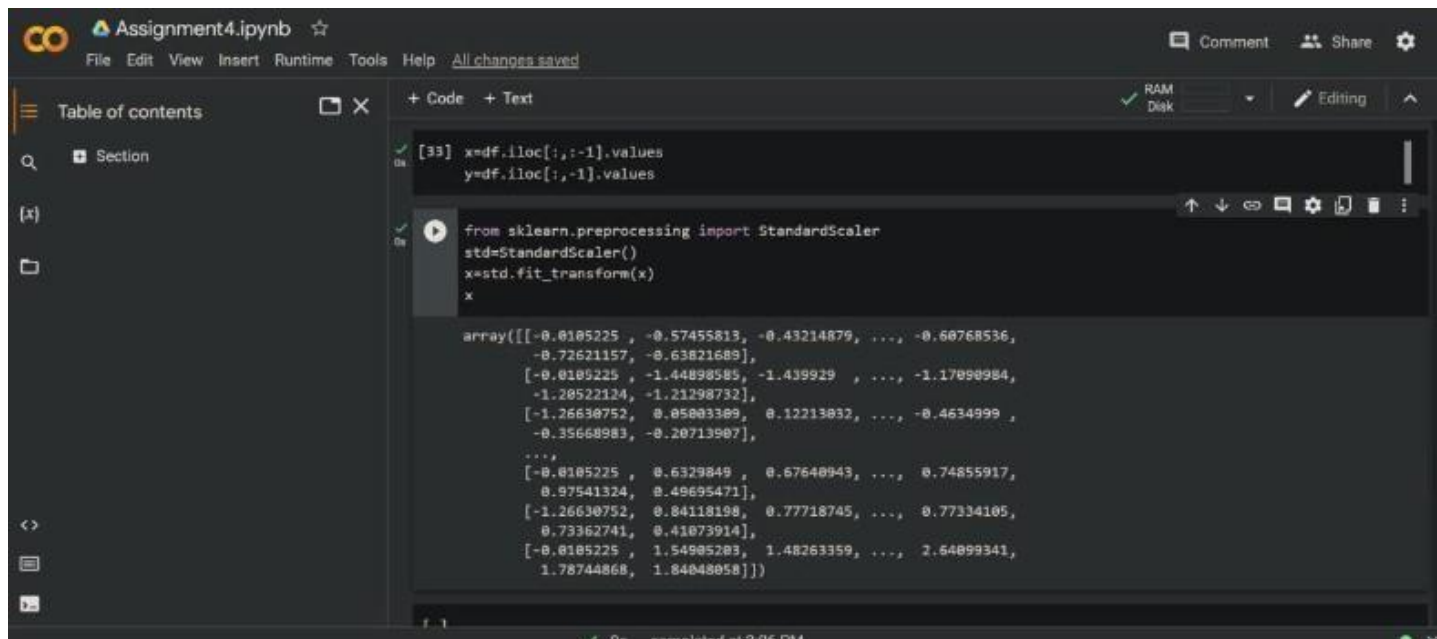
Section

(x)

[28] df['Sex'].replace({'M':1,'F':0,'I':2},inplace=True)  
df.head()

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	1	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15
1	1	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	1	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10
4	2	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7

Split the data into dependent and independent variables, Scale the independent variable



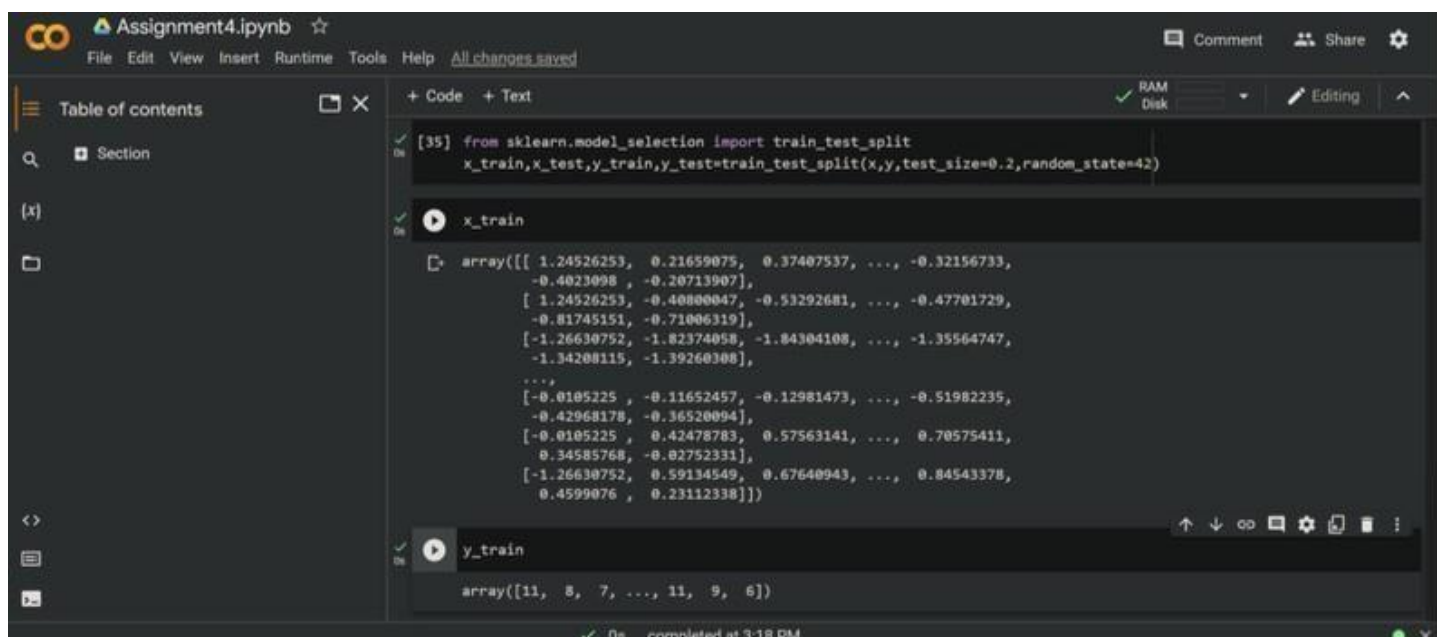
This screenshot shows a Jupyter Notebook cell where data is scaled using `StandardScaler`. The code imports `StandardScaler` from `sklearn.preprocessing`, creates an instance, and applies `fit_transform` to the data. The output is a NumPy array of scaled features.

```
[33] x=df.iloc[:,1].values
     y=df.iloc[:,2].values

from sklearn.preprocessing import StandardScaler
std=StandardScaler()
x=std.fit_transform(x)
x

array([[ -0.0105225,  -0.57455813,  -0.43214879, ...,  -0.60768536,
        -0.72621157,  -0.63821689],
       [ -0.0105225,  -1.44898585,  -1.439929, ...,  -1.17090984,
        -1.20522124,  -1.21298732],
       [ -1.26630752,  0.05003309,  0.12213032, ...,  -0.4634999,
        -0.35668983,  -0.20713907],
       ...,
       [ -0.0105225,  0.6329849,  0.67640943, ...,  0.74855917,
        0.97541324,  0.49695471],
       [ -1.26630752,  0.84118198,  0.77718745, ...,  0.77334105,
        0.73362741,  0.41073914],
       [ -0.0105225,  1.54905203,  1.48263359, ...,  2.64099341,
        1.78744868,  1.84048058]])
```

Split the data into training and testing



This screenshot shows a Jupyter Notebook cell where the data is split into training and testing sets using `train_test_split` from `sklearn.model_selection`. The code splits the data with a test size of 0.2 and a random state of 42. The resulting training features (`x_train`) and training labels (`y_train`) are displayed as NumPy arrays.

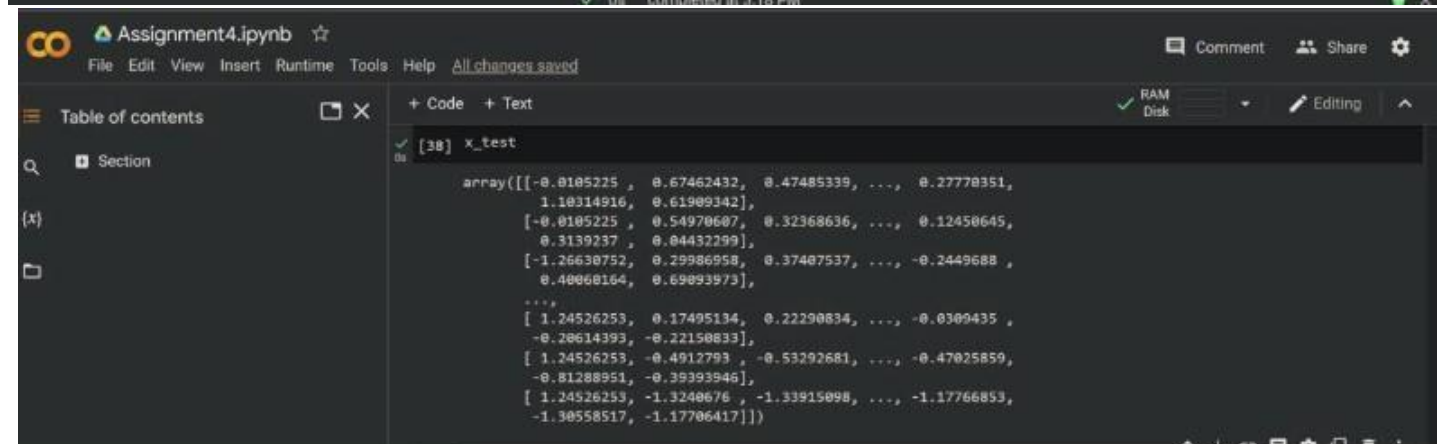
```
[35] from sklearn.model_selection import train_test_split
     x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)

x_train

array([[ 1.24526253,  0.21659075,  0.37407537, ...,  -0.32156733,
        -0.4023098,  -0.20713907],
       [ 1.24526253,  -0.40800047,  -0.53292681, ...,  -0.47701729,
        -0.81745151,  -0.71006319],
       [ -1.26630752,  -1.82374058,  -1.84304108, ...,  -1.35564747,
        -1.34208115,  -1.39260308],
       ...,
       [ -0.0105225,  -0.11652457,  -0.12981473, ...,  -0.51982235,
        -0.42968178,  -0.36520094],
       [ -0.0105225,  0.42478783,  0.57563141, ...,  0.70575411,
        0.34585768,  -0.02752331],
       [ -1.26630752,  0.59134549,  0.67640943, ...,  0.84543378,
        0.4599076,  0.23112338]])

y_train

array([11,  8,  7, ..., 11,  9,  6])
```

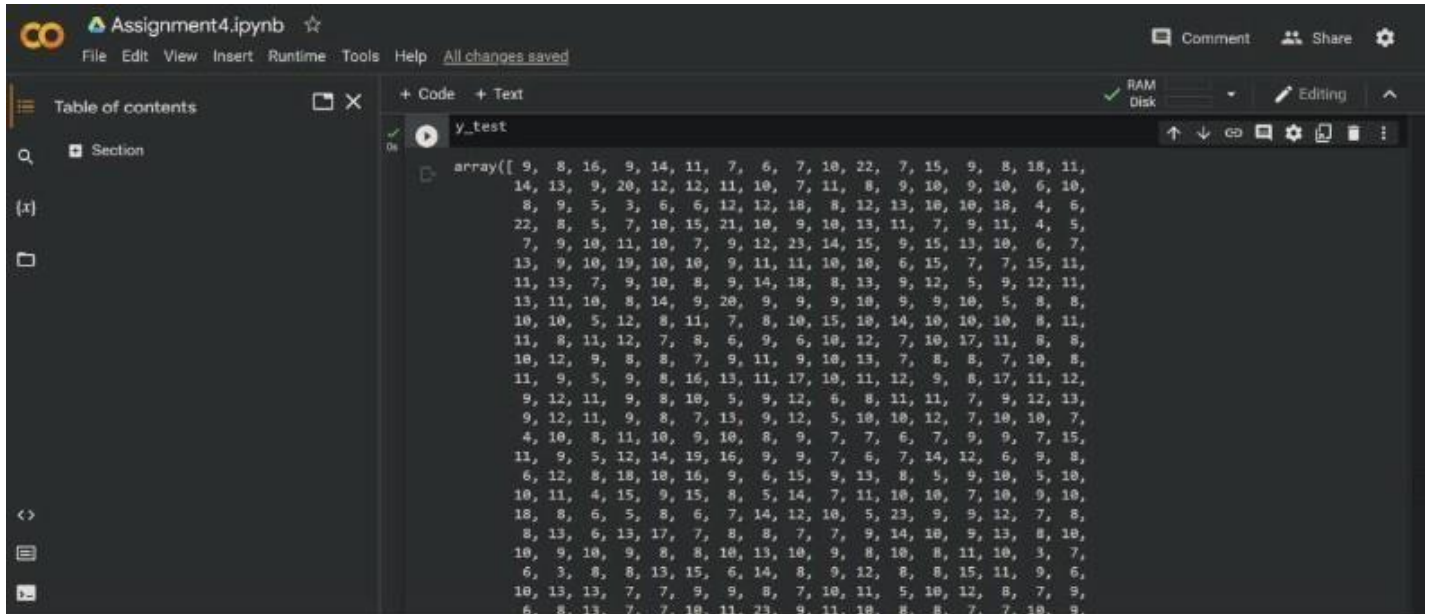


This screenshot shows a Jupyter Notebook cell displaying the test set features (`x_test`) as a NumPy array.

```
[38] x_test

array([[ -0.0105225,  0.67462432,  0.47485339, ...,  0.27770351,
         1.10314916,  0.61909342],
       [ -0.0105225,  0.54970607,  0.32368636, ...,  0.12450645,
         0.3139237,  0.04432299],
       [ -1.26630752,  0.29986958,  0.37407537, ...,  -0.2449688,
         0.40060164,  0.69093973],
       ...,
       [ 1.24526253,  0.17495134,  0.22290834, ...,  -0.0309435,
        -0.20614393,  -0.22150833],
       [ 1.24526253,  -0.4912793,  -0.53292681, ...,  -0.47025859,
        -0.81288951,  -0.39393946],
       [ 1.24526253,  -1.3240676,  -1.33915098, ...,  -1.17766853,
        -1.30558517,  -1.17706417]])
```





Assignment4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

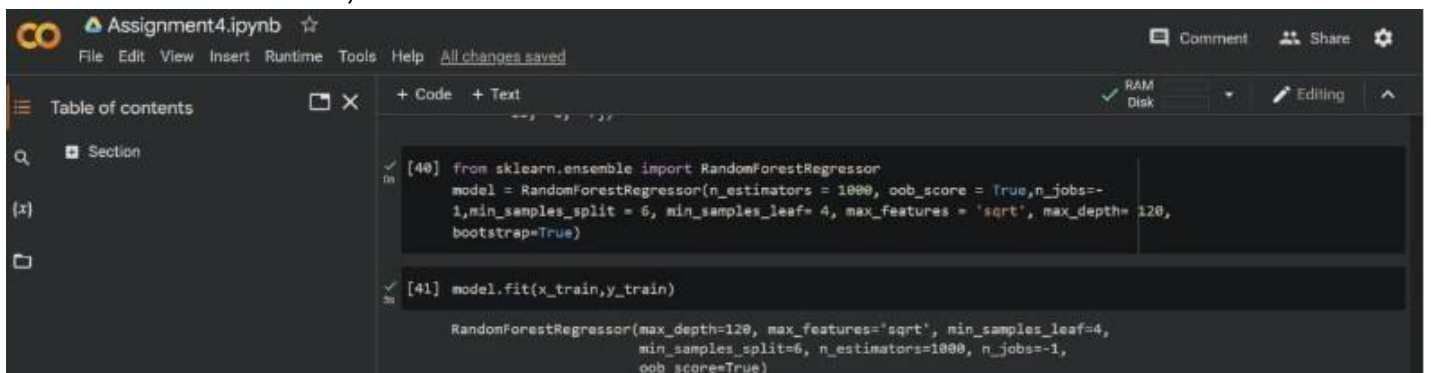
Table of contents

Section

y\_test

```
array([ 9,  8, 16,  9, 14, 11,  7,  6,  7, 10, 22,  7, 15,  9,  8, 18, 11,
       14, 13,  9, 20, 12, 12, 11, 10,  7, 11,  8,  9, 10,  9, 10,  6, 10,
        8,  9,  5,  3,  6,  6, 12, 12, 18,  8, 12, 13, 10, 10, 18,  4,  6,
       22,  8,  5,  7, 18, 15, 21, 10,  9, 10, 13, 11,  7,  9, 11,  4,  5,
        7,  9, 10, 11, 10,  7,  9, 12, 23, 14, 15,  9, 15, 13, 10,  6,  7,
       13,  9, 10, 19, 10, 10,  9, 11, 11, 10, 10,  6, 15,  7,  7, 15, 11,
       11, 13,  7,  9, 10,  8,  9, 14, 18,  8, 13,  9, 12,  5,  9, 12, 11,
       13, 11, 10,  8, 14,  9, 20,  9,  9,  9, 10,  9,  9, 10,  5,  8,  8,
       10, 10,  5, 12,  8, 11,  7,  8, 10, 15, 10, 14, 10, 10, 10,  8, 11,
       11,  8, 11, 12,  7,  8,  6,  9,  6, 10, 12,  7, 10, 17, 11,  8,  8,
       10, 12,  9,  8,  8,  7,  9, 11,  9, 10, 13,  7,  8,  8,  7, 10,  8,
       11,  9,  5,  9,  8, 16, 13, 11, 17, 10, 11, 12,  9,  8, 17, 11, 12,
        9, 12, 11,  9,  8, 10,  3,  9, 12,  6,  8, 11, 11,  7,  9, 12, 13,
        9, 12, 11,  9,  8,  7, 13,  9, 12,  5, 10, 10, 12,  7, 10, 10,  7,
        4, 10,  8, 11, 10,  9, 10,  8,  9,  7,  7,  6,  7,  9,  9,  7, 15,
       11,  9,  5, 12, 14, 19, 16,  9,  9,  7,  6,  7, 14, 12,  6,  9,  8,
        6, 12,  8, 18, 10, 16,  9,  6, 15,  9, 13,  8,  5,  9, 10,  5, 10,
       10, 11,  4, 15,  9, 15,  8,  5, 14,  7, 11, 10, 10,  7, 10,  9, 10,
       18,  8,  6,  5,  8,  6,  7, 14, 12, 10,  5, 23,  9,  9, 12,  7,  8,
        8, 13,  6, 13, 17,  7,  8,  8,  7,  7,  9, 14, 10,  9, 13,  8, 10,
       10,  9, 10,  9,  8,  8, 10, 13, 10,  9,  8, 10,  8, 11, 10,  3,  7,
        6,  3,  8,  8, 13, 15,  6, 14,  8,  9, 12,  8,  8, 15, 11,  9,  6,
       10, 13, 13,  7,  7,  9,  9,  8,  7, 10, 11,  5, 10, 12,  8,  7,  9,
        6,  8, 13,  7,  7, 10, 11, 23,  9, 11, 10,  8,  8,  7,  7, 10,  9,
```

## Build the model, Train the Model Test the Model



Assignment4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

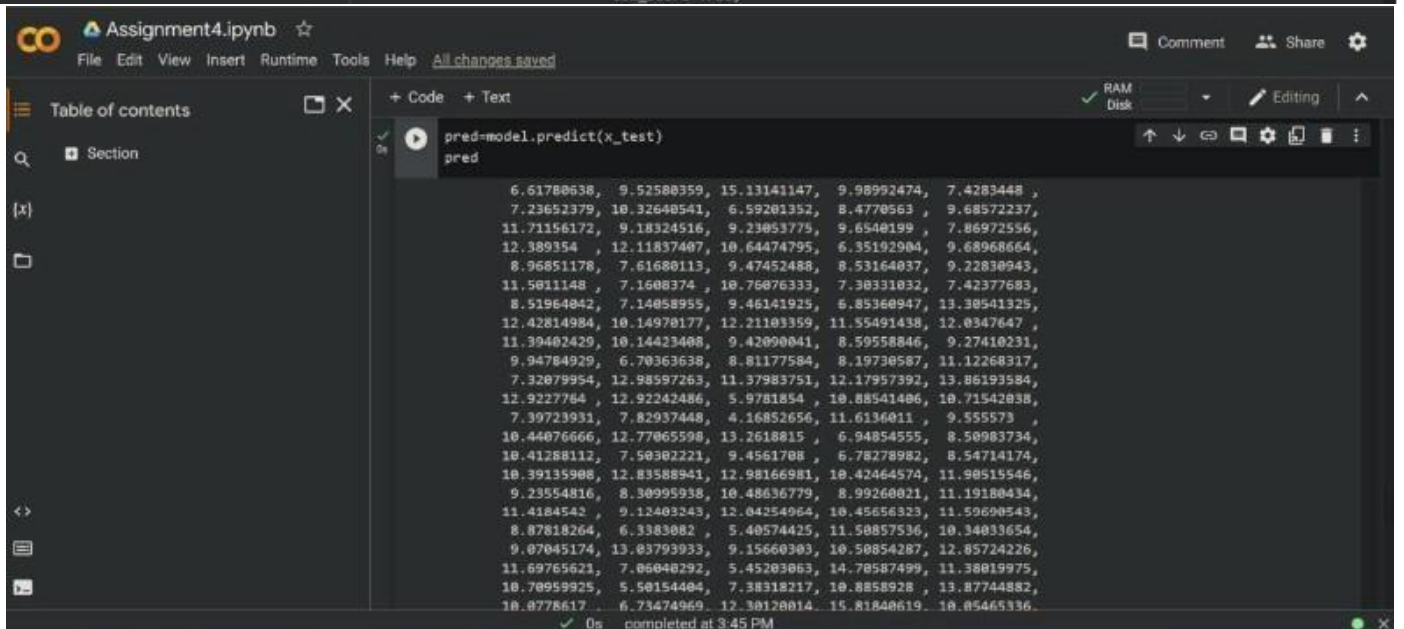
Table of contents

Section

```
[40] from sklearn.ensemble import RandomForestRegressor
      model = RandomForestRegressor(n_estimators = 1000, oob_score = True, n_jobs=-1, min_samples_split = 6, min_samples_leaf = 4, max_features = 'sqrt', max_depth= 120, bootstrap=True)

[41] model.fit(x_train,y_train)
```

RandomForestRegressor(max\_depth=120, max\_features='sqrt', min\_samples\_leaf=4, min\_samples\_split=6, n\_estimators=1000, n\_jobs=-1, oob\_score=True)



Assignment4.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Table of contents


Section

```
pred=model.predict(x_test)
pred
```

```
6.61780638, 9.52580359, 15.13141147, 9.98992474, 7.4283448 ,
7.23652379, 10.32640541, 6.59201352, 8.4770563 , 9.68572237,
11.71156172, 9.18324516, 9.23053775, 9.6540199 , 7.86972556,
12.389354 , 12.11837407, 10.64474795, 6.35192904, 9.68968664,
8.96851178, 7.61680113, 9.47452488, 8.53164037, 9.22830943,
11.5011148 , 7.1608374 , 10.76076333, 7.30331032, 7.42377683,
8.51964042, 7.14058955, 9.46141925, 6.85360947, 13.30541325,
12.42814984, 10.14970177, 12.21103359, 11.55491438, 12.0347647 ,
11.39402429, 10.14423408, 9.42090041, 8.59558846, 9.27410231,
9.94784929, 6.70363638, 8.81177584, 8.19730587, 11.12268317,
7.32079954, 12.98597263, 11.37983751, 12.17957392, 13.86193584,
12.9227764 , 12.92242486, 5.9781854 , 10.88541406, 10.71542038,
7.39723931, 7.82937448, 4.16852656, 11.6136011 , 9.555573 ,
10.44076666, 12.77065598, 13.2618815 , 6.94854555, 8.50903734,
10.41288112, 7.50302221, 9.4561708 , 6.78278982, 8.54714174,
10.39135908, 12.83588941, 12.98166981, 10.42464574, 11.90515546,
9.23554816, 8.30995938, 10.48636779, 8.99260021, 11.19180434,
11.4184542 , 9.12403243, 12.04254064, 10.45656323, 11.50600543,
8.87818264, 6.3383082 , 5.40574425, 11.50857536, 10.34033654,
9.07045174, 13.03793933, 9.15660303, 10.50854287, 12.85724226,
11.69765621, 7.06040292, 5.45203063, 14.70587499, 11.38019975,
10.70959925, 5.50154404, 7.38318217, 10.8858928 , 13.87744882,
10.8778617 , 6.73474969, 12.30120014, 15.81840619, 10.05465336,
```

completed at 3:45 PM

## Measure the performance using metrics

 Assignment4.ipynb ☆

File Edit View Insert Runtime Tools Help All changes saved

Table of contents

Section

(x)

+ Code + Text

RAM Disk

Editing

0s

```
from sklearn.metrics import r2_score
acc=r2_score(y_test,pred)
acc
```

0.5565430591634158