

ASSIGNMENT-3

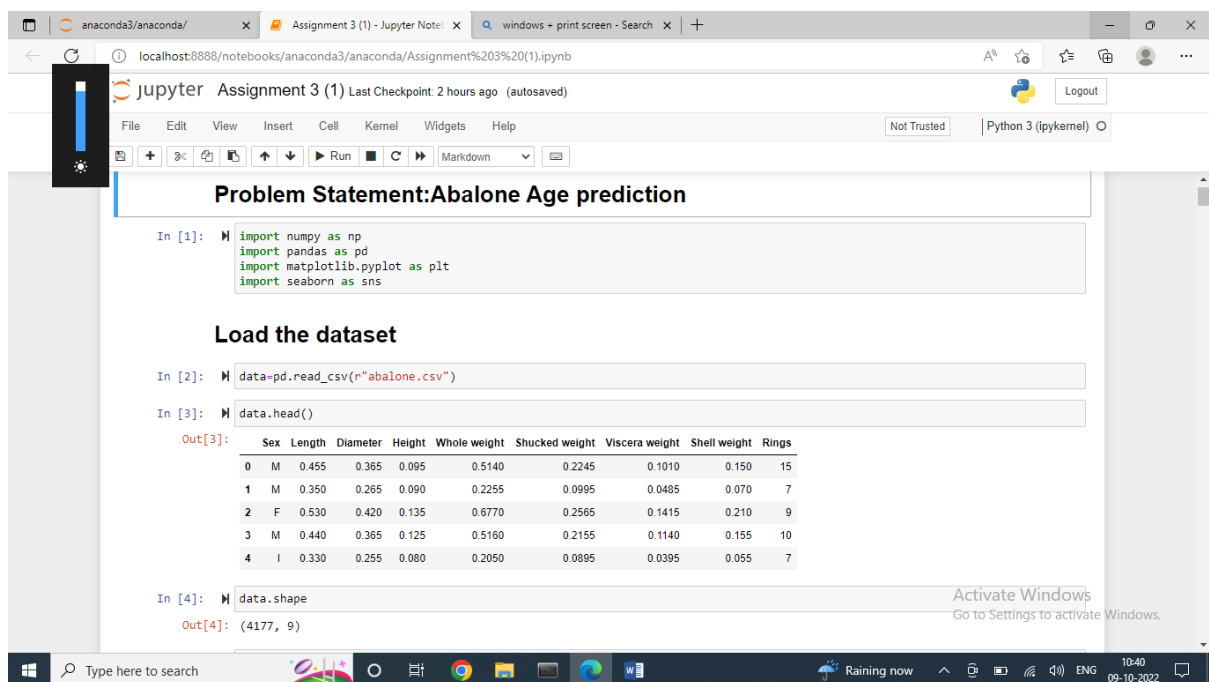
Applied Data Science

Assignment Date:08 October 2022

Student name :Ms.Varshini

Student roll number:721719104089

Maximum marks:2 marks



The screenshot displays a Jupyter Notebook titled "Assignment 3 (1)" running on a local host. The notebook contains the following content:

Problem Statement: Abalone Age prediction

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

Load the dataset

```
In [2]: data=pd.read_csv(r"abalone.csv")
```

```
In [3]: data.head()
```

Out[3]:

	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

```
In [4]: data.shape
```

Out[4]: (4177, 9)

The interface also shows a "Not Trusted" warning and a "Python 3 (ipykernel)" environment indicator. The Windows taskbar at the bottom shows the time as 10:40 on 09-10-2022.

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows + print screen - Search x +

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [5]: data.size
Out[5]: 37593

In [6]: data.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 4177 entries, 0 to 4176  
Data columns (total 9 columns):  
#   Column              Non-Null Count  Dtype  
---  ---  
0   Sex                  4177 non-null   object  
1   Length               4177 non-null   float64  
2   Diameter             4177 non-null   float64  
3   Height               4177 non-null   float64  
4   Whole weight         4177 non-null   float64  
5   Shucked weight       4177 non-null   float64  
6   Viscera weight       4177 non-null   float64  
7   Shell weight         4177 non-null   float64  
8   Rings               4177 non-null   int64  
dtypes: float64(7), int64(1), object(1)  
memory usage: 293.8+ KB
```

Data Analysis & Visualization

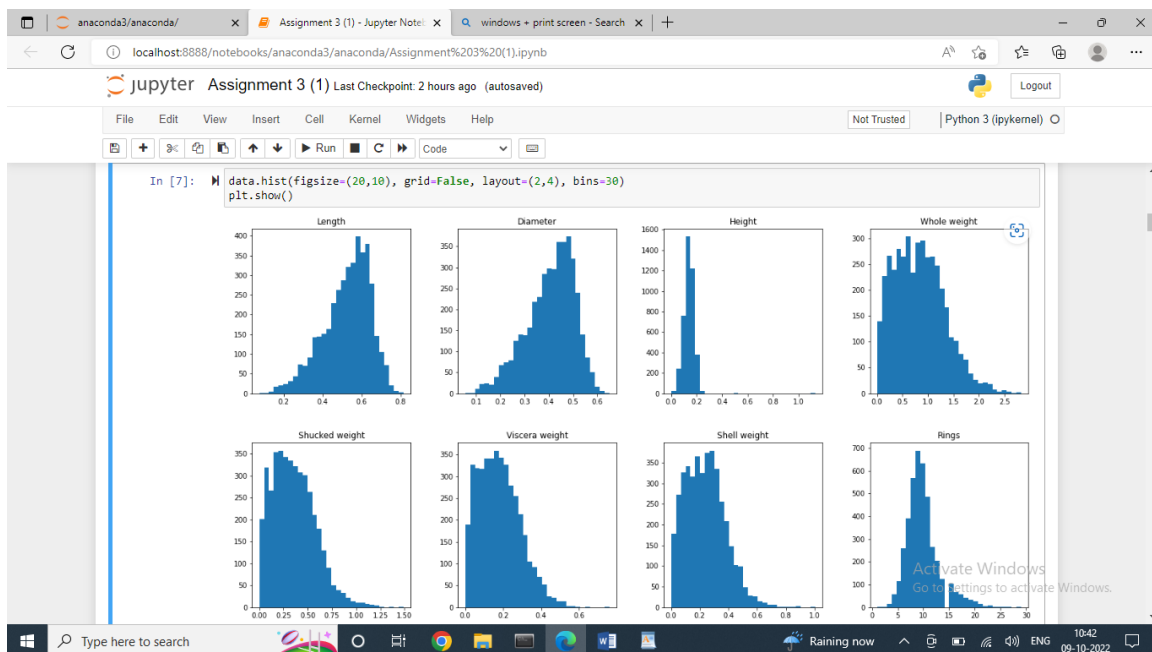
Univariate Analysis

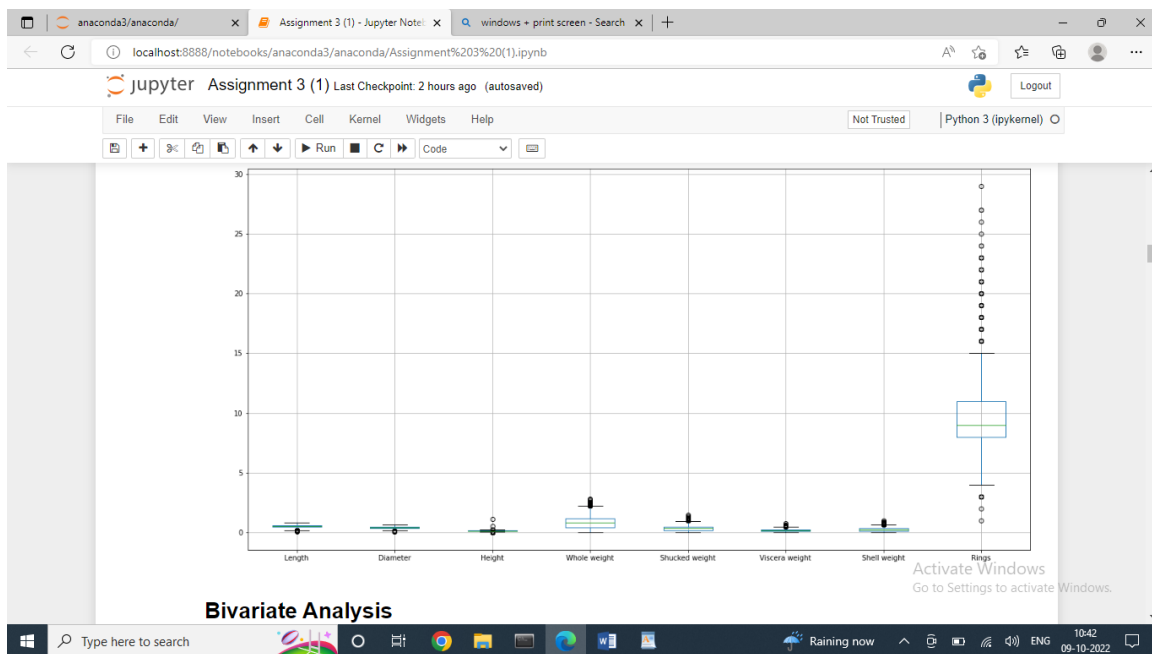
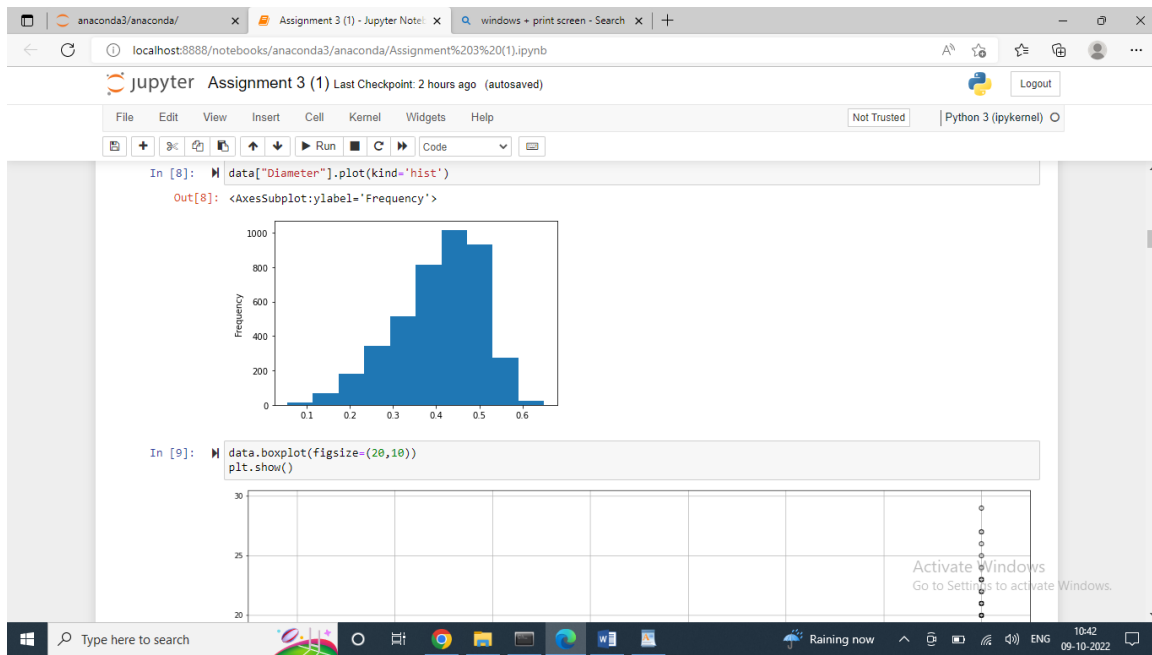
Activate Windows
Go to Settings to activate Windows.

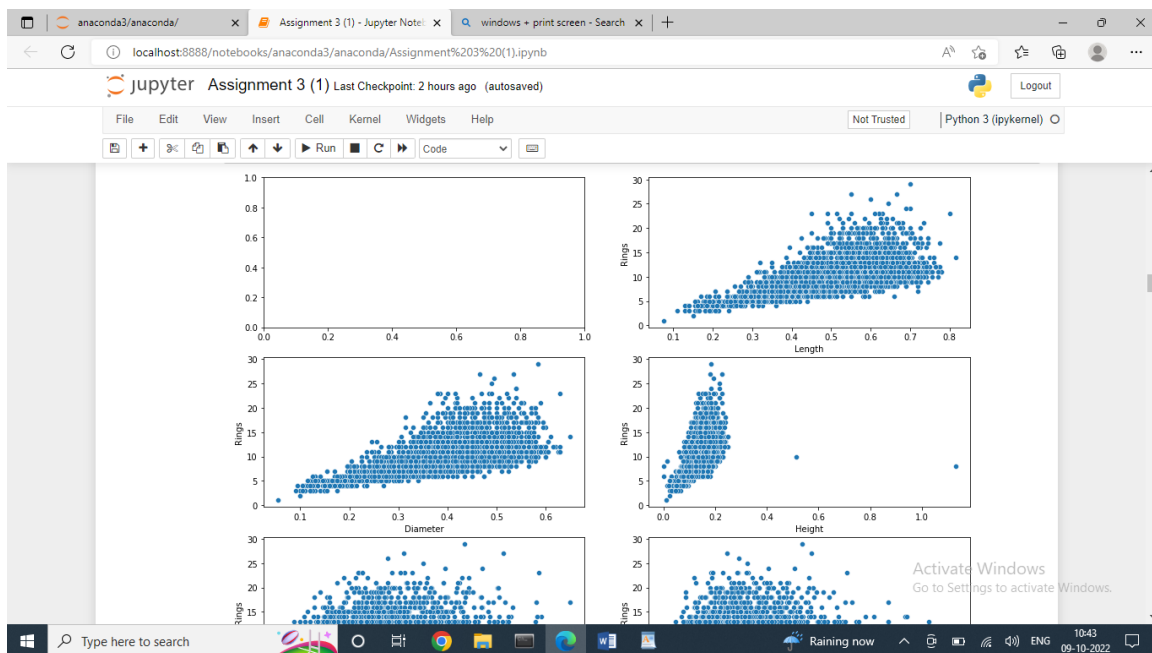
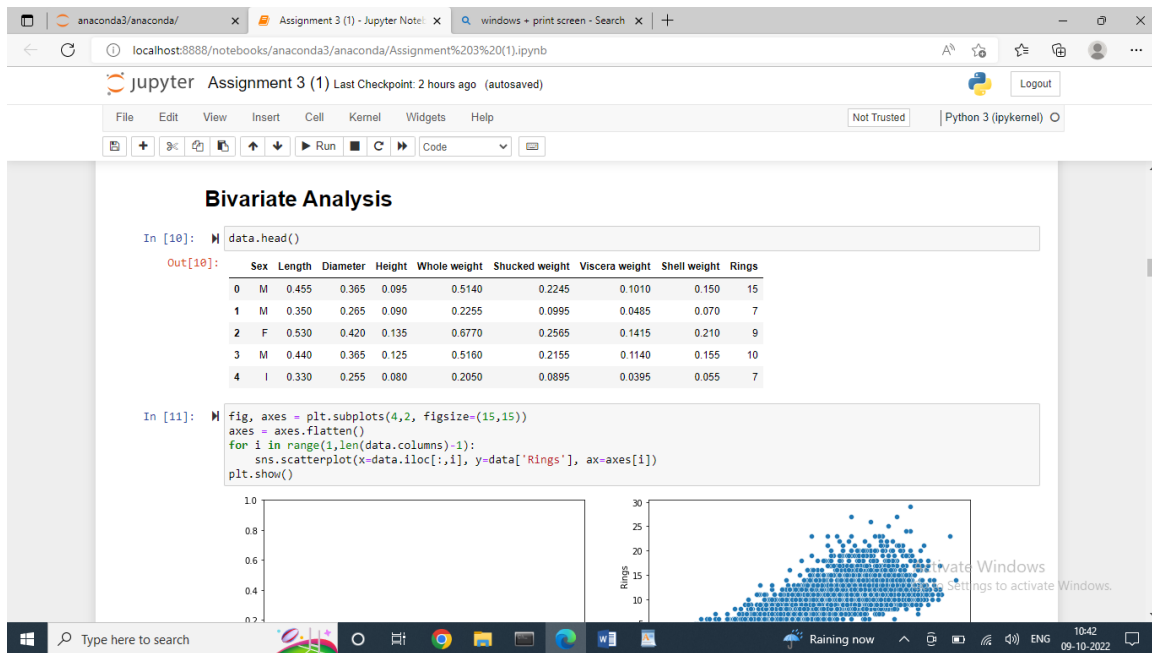
Type here to search

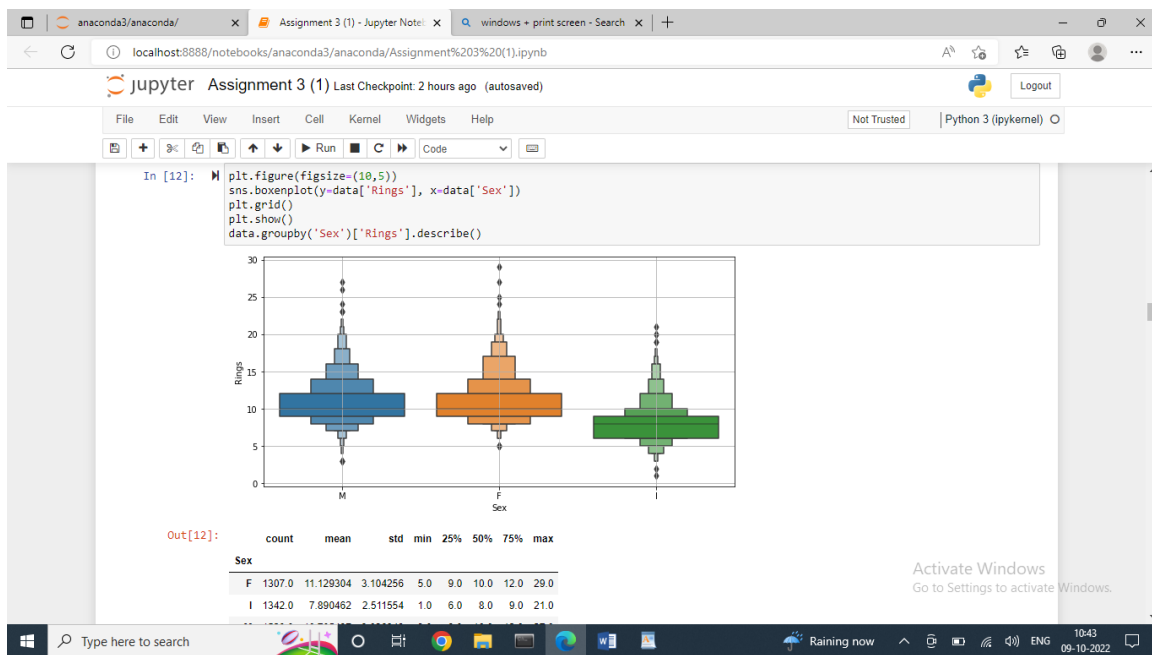
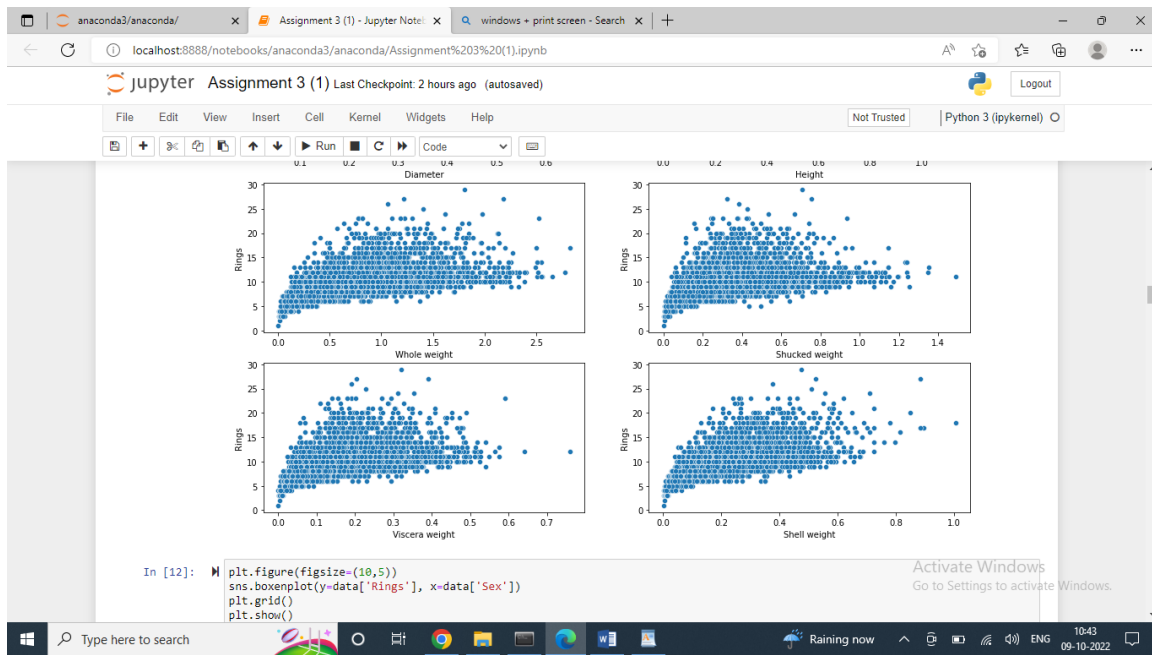
Raining now

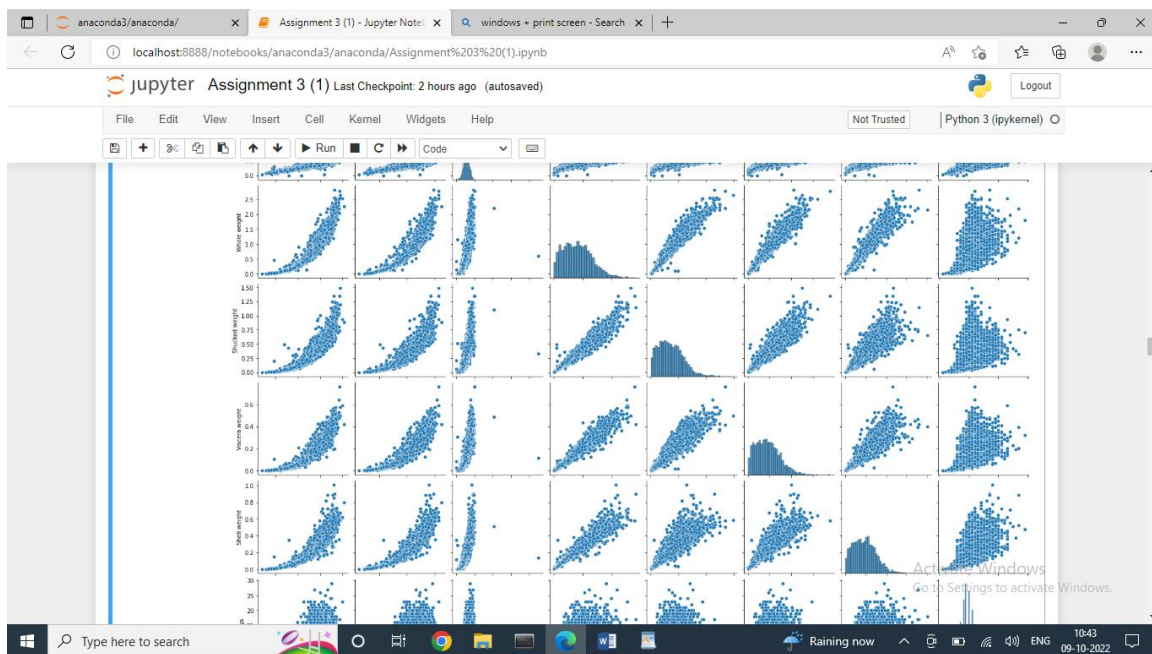
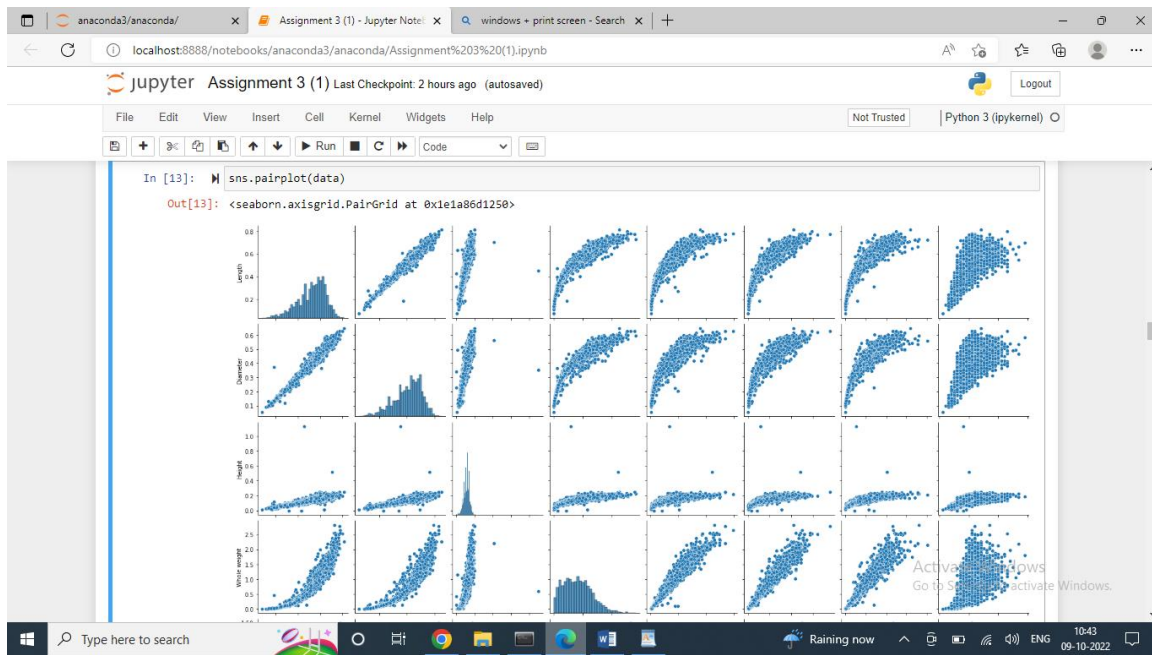
10:41
09-10-2022











anaconda3/anaconda/ Assignment 3 (1) - Jupyter Note: windows - print screen - Search

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Descriptive statistics

```
In [14]: data.describe()
```

```
Out[14]:
```

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
count	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000
mean	0.523992	0.407881	0.139516	0.028742	0.359367	0.180594	0.238831	9.933684
std	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0.139203	3.224169
min	0.075000	0.055000	0.000000	0.002000	0.001000	0.000500	0.001500	1.000000
25%	0.450000	0.350000	0.115000	0.441500	0.186000	0.093500	0.130000	8.000000
50%	0.545000	0.425000	0.140000	0.799500	0.336000	0.171000	0.234000	9.000000
75%	0.615000	0.480000	0.165000	1.153000	0.502000	0.253000	0.329000	11.000000
max	0.815000	0.650000	1.130000	2.825500	1.488000	0.780000	1.005000	29.000000

Handle The Missing Values

```
In [15]: data.isnull().any()
```

```
Out[15]: Sex          False
Length        False
Diameter       False
Height         False
Whole weight   False
Shucked weight False
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:44 09-10-2022

anaconda3/anaconda/ Assignment 3 (1) - Jupyter Note: windows - print screen - Search

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [16]: data.isnull().sum()
```

```
Out[16]: Sex          0
Length        0
Diameter       0
Height         0
Whole weight   0
Shucked weight 0
Viscera weight 0
Shell weight   0
Rings          0
dtype: int64
```

Find the outliers

```
In [17]: data.skew()
```

```
Out[17]: Length          -0.639873
Diameter         -0.609198
Height           3.128817
Whole weight      0.530959
Shucked weight    0.719098
Viscera weight    0.591852
Shell weight      0.620927
Rings             1.114102
```

C:\Users\admin\AppData\Local\Temp\ipykernel_11300\1188251951.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.
data.skew()

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:44 09-10-2022

anaconda3/anaconda/ Assignment 3 (1) - Jupyter Note: windows - print screen - Search

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Find the outliers

```
In [17]: data.skew()
```

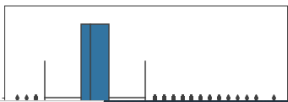
C:\Users\admin\AppData\Local\Temp\ipykernel_11300\1188251951.py:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is deprecated; in a future version this will raise TypeError. Select only valid columns before calling the reduction.

```
data.skew()
```

```
Out[17]: Length      -0.639873
Diameter    -0.609198
Height       3.128817
Whole weight  0.530959
Shucked weight 0.719098
Viscera weight 0.591852
Shell weight  0.620927
Rings        1.114102
dtype: float64
```

```
In [18]: sns.boxplot(x=data['Rings'],data=data)
```

```
Out[18]: <AxesSubplot:xlabel='Rings'>
```



Activate Windows
Go to Settings to activate Windows.

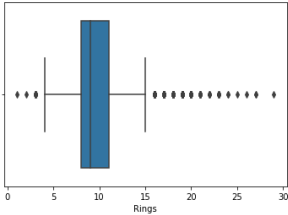
Type here to search Raining now 10:44 09-10-2022

anaconda3/anaconda/ Assignment 3 (1) - Jupyter Note: windows - print screen - Search

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

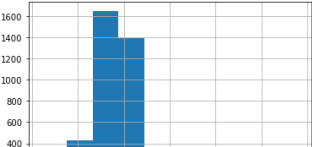
jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)



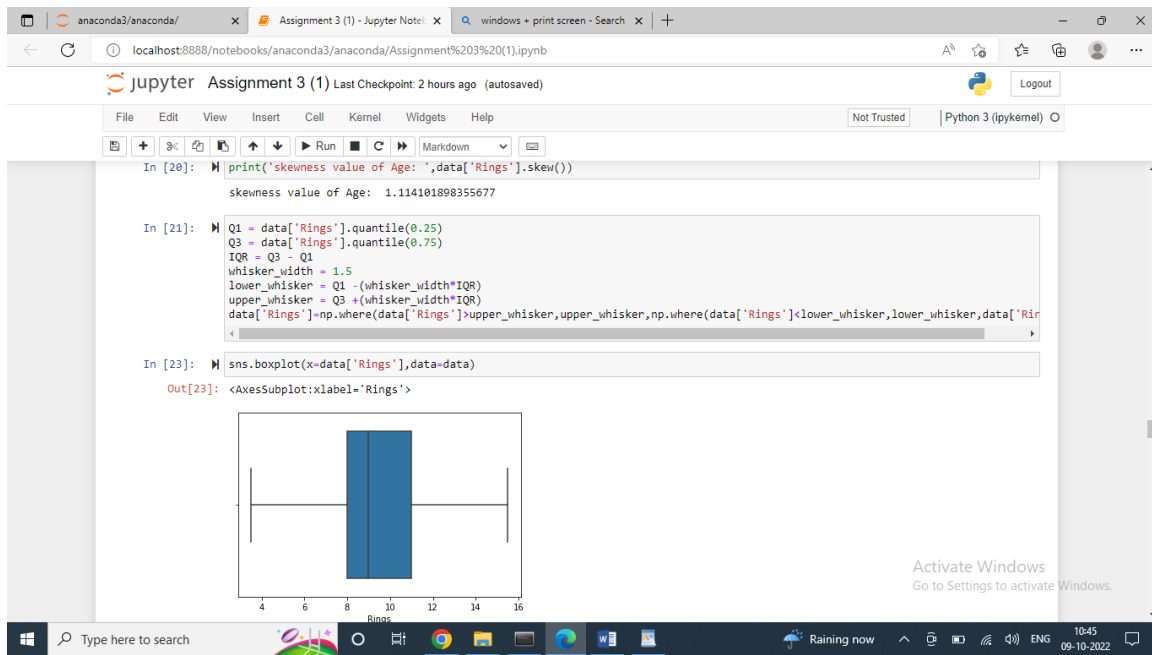
```
In [19]: data['Rings'].hist()
```

```
Out[19]: <AxesSubplot:>
```



Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:44 09-10-2022



anaconda3/anaconda/ Assignment 3 (1) - Jupyter Note: windows - print screen - Search

localhost:8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Categorical encoding

```
In [24]: data.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4177 entries, 0 to 4176
Data columns (total 9 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Sex                  4177 non-null   object  
1   Length               4177 non-null   float64 
2   Diameter             4177 non-null   float64 
3   Height              4177 non-null   float64 
4   Whole weight         4177 non-null   float64 
5   Shucked weight       4177 non-null   float64 
6   Viscera weight        4177 non-null   float64 
7   Shell weight         4177 non-null   float64 
8   Rings                4177 non-null   float64 
dtypes: float64(8), object(1)
memory usage: 293.8+ KB
```

```
In [25]: from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
data['Sex']=le.fit_transform(data['Sex'])
```

```
In [26]: data.head()
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:45 09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows + print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [25]: from sklearn.preprocessing import LabelEncoder
         le=LabelEncoder()
         data['Sex']=le.fit_transform(data['Sex'])

In [26]: data.head()
```

Out[26]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15.0
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7.0
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9.0
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10.0
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7.0

```
In [28]: data["Sex"].unique()
Out[28]: array([2, 0, 1])
```

Split the data into training and testing

```
In [29]: data.head(5)
```

Out[29]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15.0

Activate Windows
Go to Settings to activate Windows.

Type here to search

Raining now

10:45
09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows + print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Split the data into training and testing

```
In [29]: data.head(5)
```

Out[29]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010	0.150	15.0
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.070	7.0
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9.0
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.155	10.0
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.055	7.0

```
In [30]: x=data.iloc[:,0:7]
         y=data.iloc[:, -1]
```

```
In [31]: x
```

Out[31]:

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight
0	2	0.455	0.365	0.095	0.5140	0.2245	0.1010
1	2	0.350	0.265	0.090	0.2255	0.0995	0.0485
2	0	0.530	0.420	0.135	0.6770	0.2565	0.1415
3	2	0.440	0.365	0.125	0.5160	0.2155	0.1140
4	1	0.330	0.255	0.080	0.2050	0.0895	0.0395

Activate Windows
Go to Settings to activate Windows.

Type here to search

Raining now

10:45
09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows - print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

4177 rows x 7 columns

```
In [32]: y
Out[32]: 0    15.0
         1     7.0
         2     9.0
         3    10.0
         4     7.0
         ...
        4172    11.0
        4173    10.0
        4174     9.0
        4175    10.0
        4176    12.0
        Name: Rings, Length: 4177, dtype: float64
```

```
In [33]: y.shape
Out[33]: (4177,)
```

Scale the independent variables

```
In [36]: from sklearn.preprocessing import StandardScaler
         ss = StandardScaler()
         x_scaled = ss.fit_transform(x)
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:45 09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows - print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
Out[33]: (4177,)
```

Scale the independent variables

```
In [36]: from sklearn.preprocessing import StandardScaler
         ss = StandardScaler()
         x_scaled = ss.fit_transform(x)
```

Split the data into training and testing

```
In [39]: from sklearn.model_selection import train_test_split
         x_train, x_test, y_train, y_test = train_test_split(x_scaled, y, test_size = 0.3, random_state = 1)
```

Build the model

Training the model and testing the model

```
In [40]: import csv
         with open("abalone.csv") as csv_file:
             csv_reader = csv.reader(csv_file)
             data = pd.DataFrame([csv_reader], index = None)
             for val in list(data[1]):
                 print(val)
         ['M', '0.455', '0.365', '0.095', '0.514', '0.2245', '0.101', '0.15', '15']
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:45 09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows - print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Training and testing Module

1.Linear Regression 2.Ridge 3.Decision tree Regression 4.KNeighborsRegressor

```
In [41]: from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Ridge
from sklearn.tree import DecisionTreeRegressor
from sklearn.neighbors import KNeighborsRegressor
from sklearn.metrics import mean_squared_error, r2_score
```

1.Linear Regression

```
In [42]: lr = LinearRegression()
lr.fit(x_train, y_train)
Out[42]: LinearRegression()

In [43]: lr_test_pred = lr.predict(x_test)

In [44]: lr_test_pred
Out[44]: array([8.49722433, 7.64369859, 7.82528883, ..., 8.55677832, 9.02884473,
5.96561877])
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:46 09-10-2022

anaconda3/anaconda/ x Assignment 3 (1) - Jupyter Note: x windows - print screen - Search x +

localhost8888/notebooks/anaconda3/anaconda/Assignment%203%20(1).ipynb

jupyter Assignment 3 (1) Last Checkpoint: 2 hours ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

```
In [42]: lr = LinearRegression()
lr.fit(x_train, y_train)
Out[42]: LinearRegression()

In [43]: lr_test_pred = lr.predict(x_test)

In [44]: lr_test_pred
Out[44]: array([8.49722433, 7.64369859, 7.82528883, ..., 8.55677832, 9.02884473,
5.96561877])

In [46]: mse = mean_squared_error(y_test, lr_test_pred)
print('Mean Squared error of testing Set: %2f'%mse)
Mean Squared error of testing Set: 3.524682

In [47]: p = r2_score(y_test, lr_test_pred)
print('R2 Score of testing set:%.2f'%p)
R2 Score of testing set:0.52
```

1.Ridge

```
In [48]: ridge_mod = Ridge(alpha=0.01, normalize=True)
ridge_mod.fit(x_train, y_train)
ridge_mod.fit(x_test, y_test)
```

Activate Windows
Go to Settings to activate Windows.

Type here to search Raining now 10:46 09-10-2022

```
Out[48]: Ridge(alpha=0.01, normalize=True)

In [49]: ridge_model_pred = ridge_mod.predict(x_test)

In [50]: ridge_model_pred
Out[50]: array([8.54831033, 8.48463396, 7.96838487, ..., 8.77493484, 9.03881023,
5.83582085])

In [51]: acc = r2_score(y_test, ridge_model_pred)
print('Score of Testing Set: %2f'%acc)
Score of testing Set: 0.523227

1.Decision Tree Regression

In [52]: dt = DecisionTreeRegressor()
dt.fit(x_train, y_train)
Out[52]: DecisionTreeRegressor()

In [53]: dt_test_pred = dt.predict(x_test)

In [54]: dt_test_pred
Out[54]: array([12., 9., 10., ..., 6., 9., 4.])

In [55]: dacc = mean_squared_error(y_test, dt_test_pred)
```

```
In [55]: dacc = mean_squared_error(y_test, dt_test_pred)
print('Mean Squared Error of testing Set: %2f'%dacc)
Mean Squared Error of testing Set: 6.182217

1.KNN Regression

In [56]: knn = KNeighborsRegressor(n_neighbors = 4 )
knn.fit(x_train, y_train)
knn.fit(x_test, y_test)
Out[56]: KNeighborsRegressor(n_neighbors=4)

In [57]: knn_test_pred = knn.predict(x_test)

In [58]: knn_test_pred
Out[58]: array([ 8.75, 9.5 , 10.5 , ..., 8. , 7.5 , 5. ])

In [59]: kacc= r2_score(knn_test_pred,y_test)
print('Score of testing Set: %2f'%kacc)
Score of testing Set: 0.400555

In [60]: kmse = mean_squared_error(knn_test_pred,y_test)
print('Score of testing Set: %2f'%kmse)
Score of testing Set: 2.602460
```

