Assignment 3

Build the Regression model

Assignment date	30 September 2022
Student Name	E. Mahalakshmi
Student Roll. No	962719106017
Maximum marks	2 marks

1.Download the Dataset

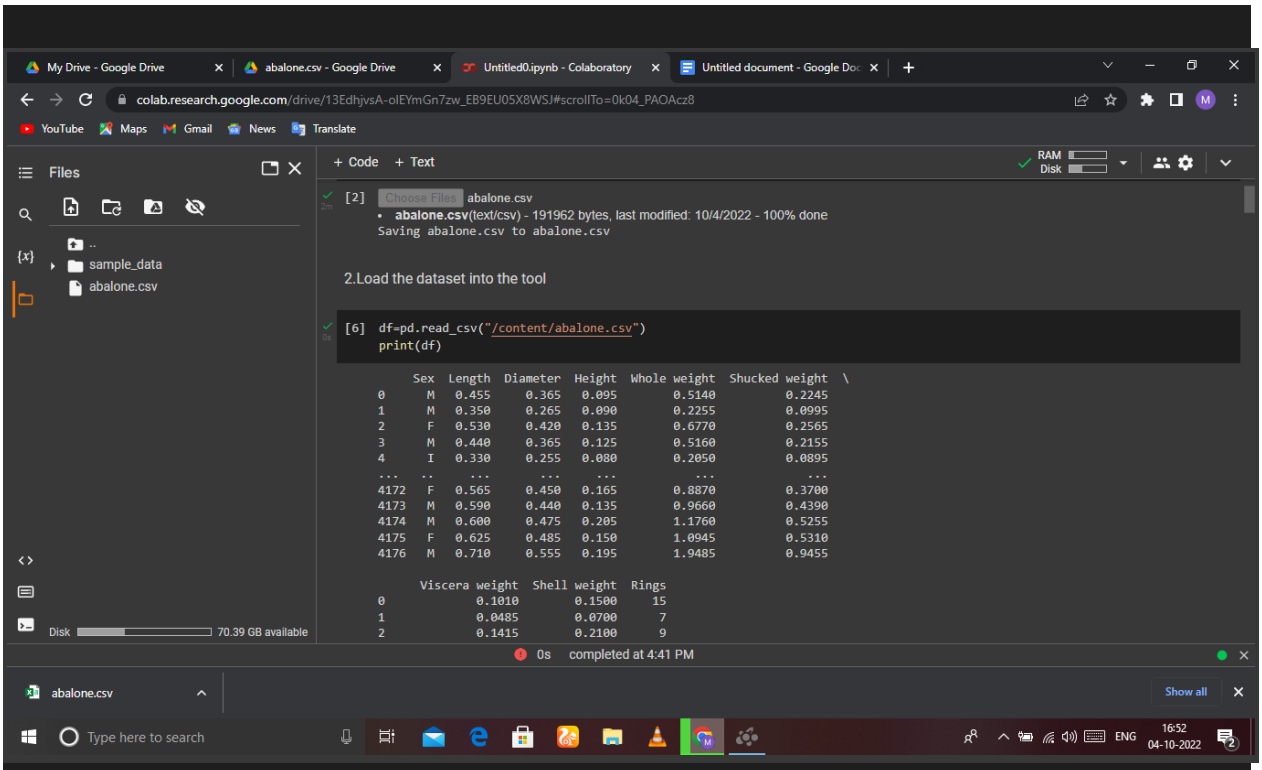
solution:

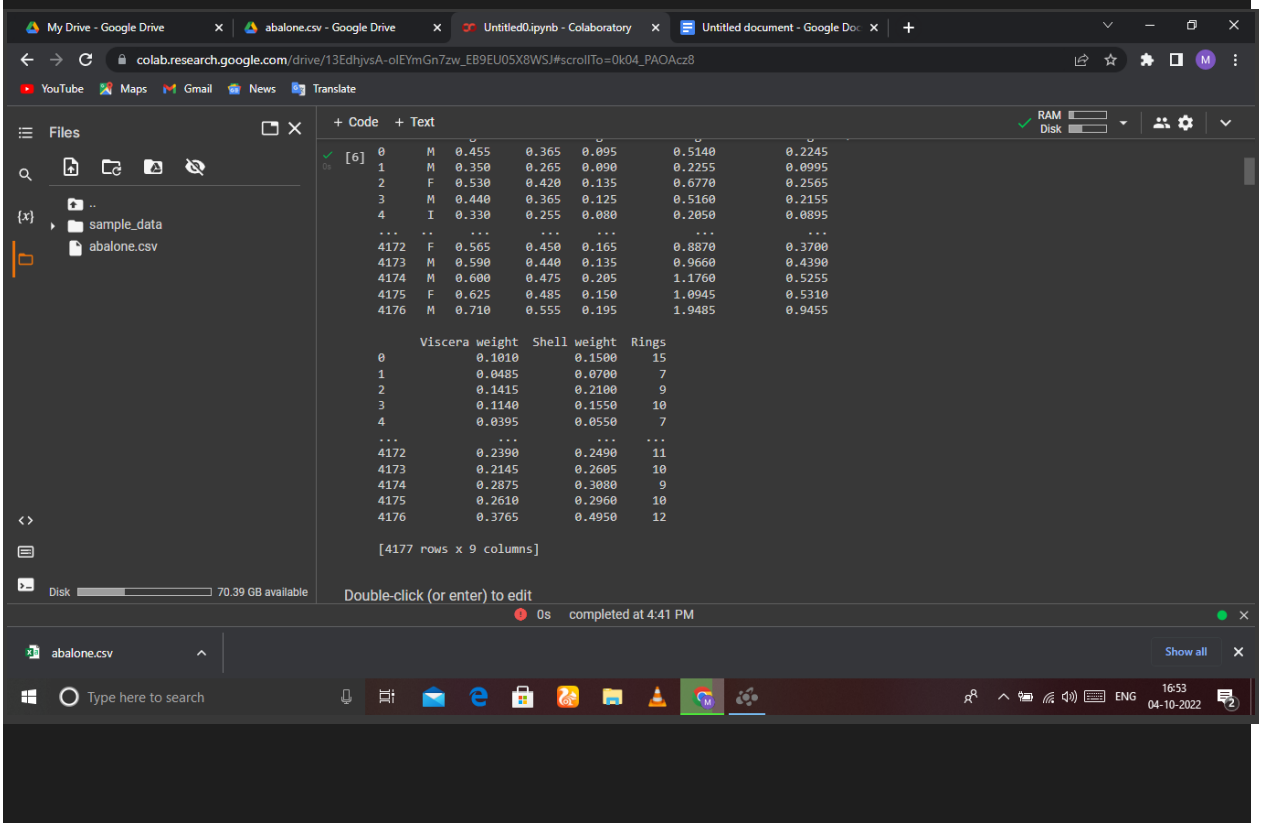
```
from google.colab import files
Upload=files.upload()
```

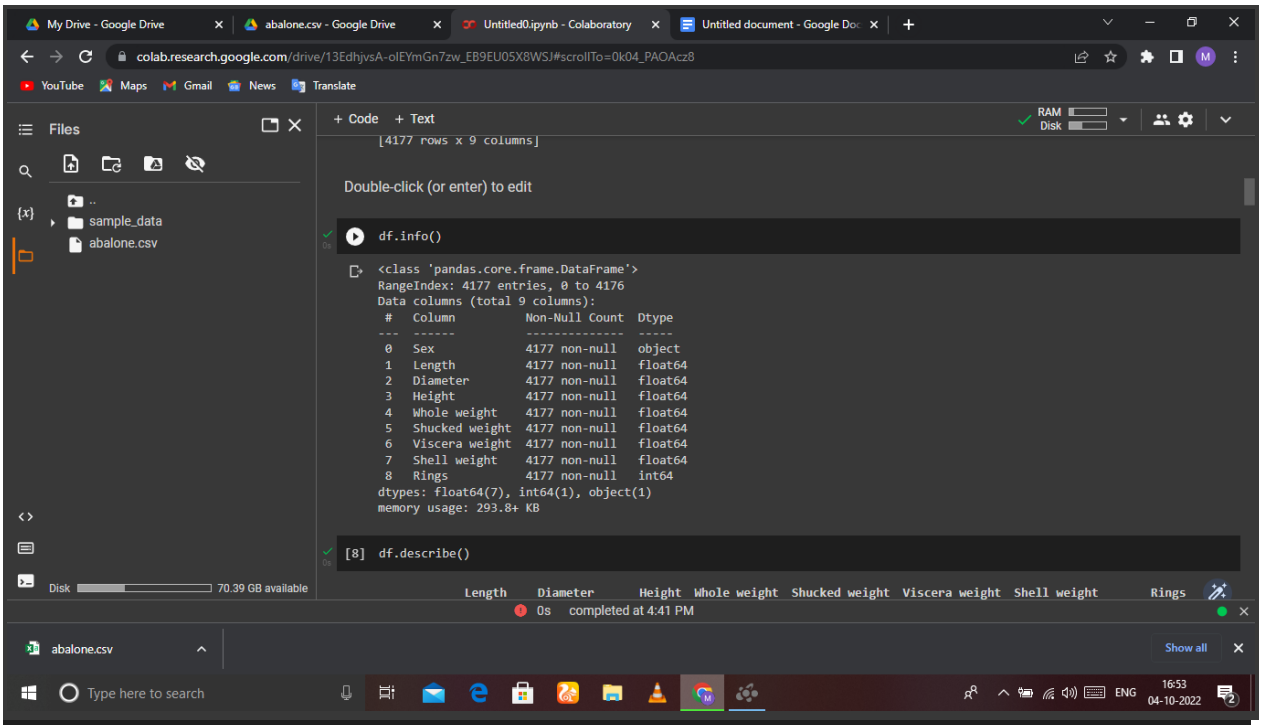
abalone.csv(text/csv) - 191962 bytes, last modified: 10/4/2022 - 100% done
Saving abalone.csv to abalone.csv

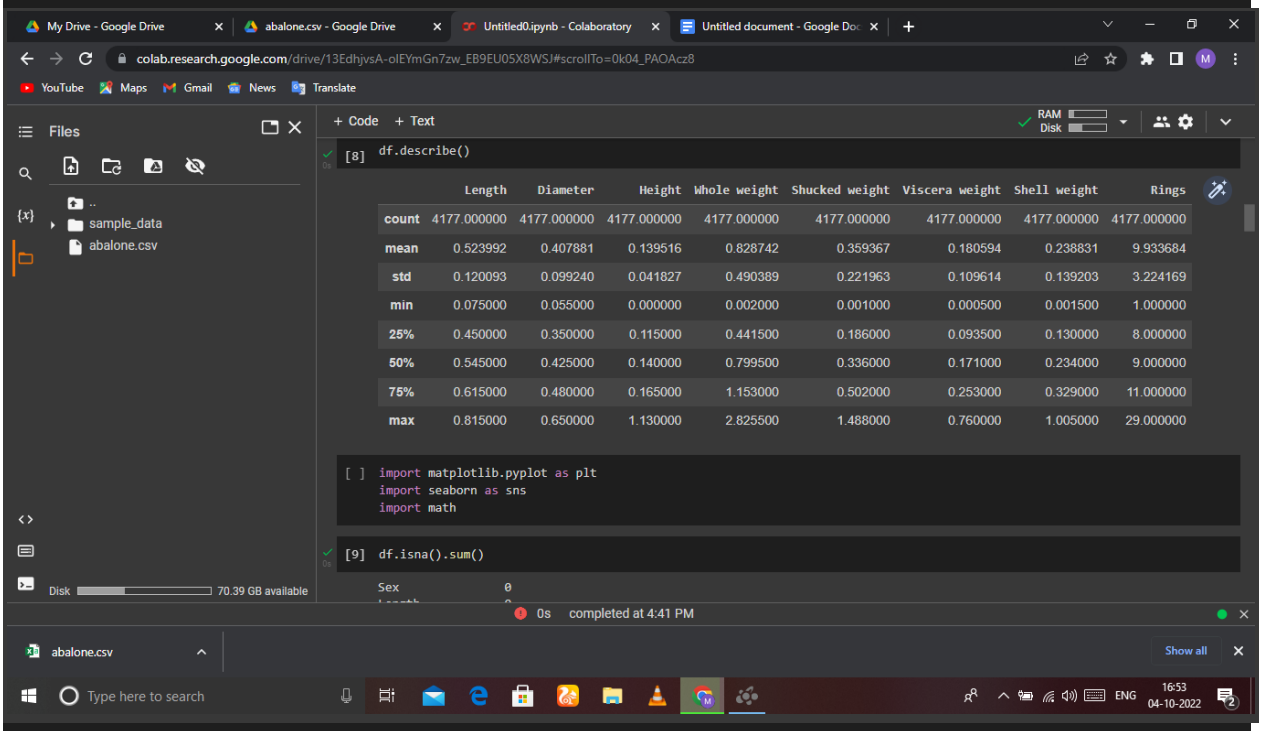
2.Load the dataset into the tool
solution:

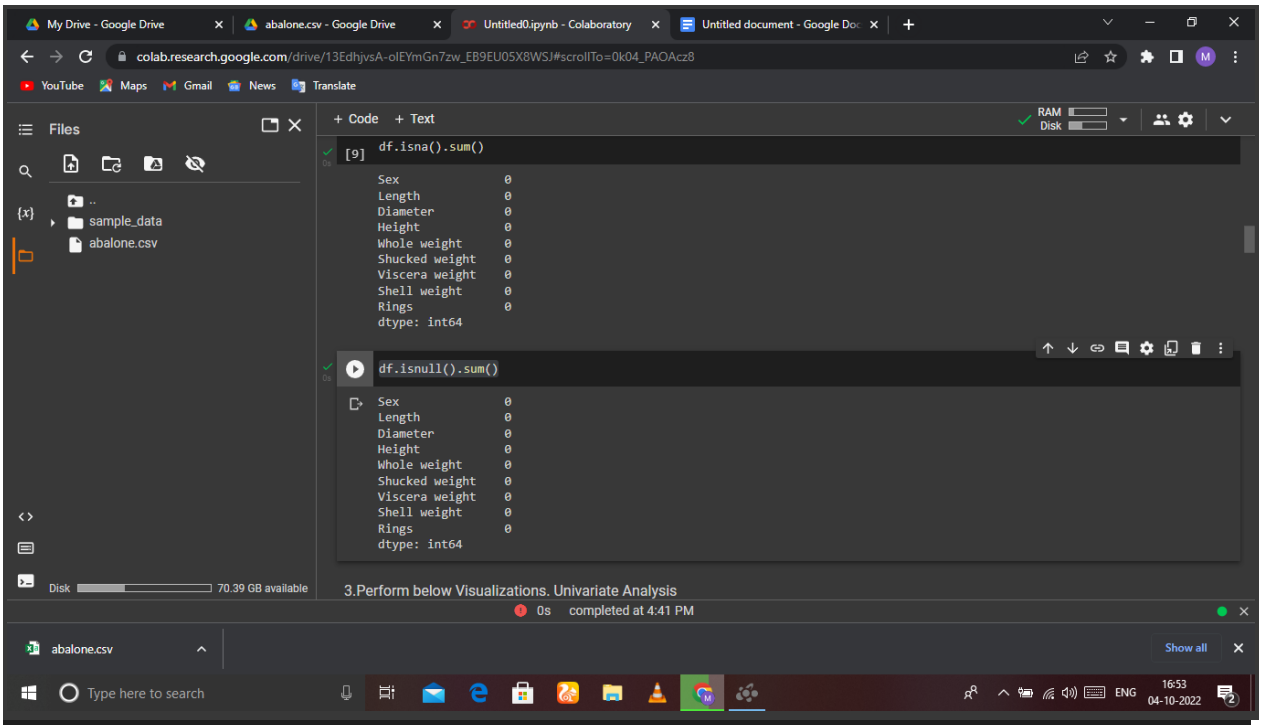
```
df=pd.read_csv("/content/abalone.csv")
print(df)
df.info()
df.describe()
import matplotlib.pyplot as plt
import seaborn as sns
import math
df.isna().sum()
df.isnull().sum()
```


My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Do
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=0k04_PAOAcz8
YouTube Maps Gmail News Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code + Text
RAM
Disk
[2] Choose Files abalone.csv
abalone.csv(text/csv) - 191962 bytes, last modified: 10/4/2022 - 100% done
Saving abalone.csv to abalone.csv
2.Load the dataset into the tool
[6] df=pd.read_csv("/content/abalone.csv")
print(df)
Sex Length Diameter Height Whole weight Shucked weight \
0 M 0.455 0.365 0.095 0.5140 0.2245
1 M 0.350 0.265 0.090 0.2255 0.0995
2 F 0.530 0.420 0.135 0.6770 0.2565
3 M 0.440 0.365 0.125 0.5160 0.2155
4 I 0.330 0.255 0.080 0.2050 0.0895
... .
4172 F 0.565 0.450 0.165 0.8870 0.3700
4173 M 0.590 0.440 0.135 0.9660 0.4390
4174 M 0.600 0.475 0.205 1.1760 0.5255
4175 F 0.625 0.485 0.150 1.0945 0.5310
4176 M 0.710 0.555 0.195 1.9485 0.9455
Viscera weight Shell weight Rings
0 0.1010 0.1500 15
1 0.0485 0.0700 7
2 0.1415 0.2100 9
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:52
04-10-2022


My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Do
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=0k04_PAOAcz8
YouTube Maps Gmail News Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code + Text
RAM
Disk
[6] 0 M 0.455 0.365 0.095 0.5140 0.2245
1 M 0.350 0.265 0.090 0.2255 0.0995
2 F 0.530 0.420 0.135 0.6770 0.2565
3 M 0.440 0.365 0.125 0.5160 0.2155
4 I 0.330 0.255 0.080 0.2050 0.0895
... .
4172 F 0.565 0.450 0.165 0.8870 0.3700
4173 M 0.590 0.440 0.135 0.9660 0.4390
4174 M 0.600 0.475 0.205 1.1760 0.5255
4175 F 0.625 0.485 0.150 1.0945 0.5310
4176 M 0.710 0.555 0.195 1.9485 0.9455
Viscera weight Shell weight Rings
0 0.1010 0.1500 15
1 0.0485 0.0700 7
2 0.1415 0.2100 9
3 0.1140 0.1550 10
4 0.0395 0.0550 7
... ...
4172 0.2390 0.2490 11
4173 0.2145 0.2605 10
4174 0.2875 0.3080 9
4175 0.2610 0.2960 10
4176 0.3765 0.4950 12
[4177 rows x 9 columns]
Double-click (or enter) to edit
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:53
04-10-2022

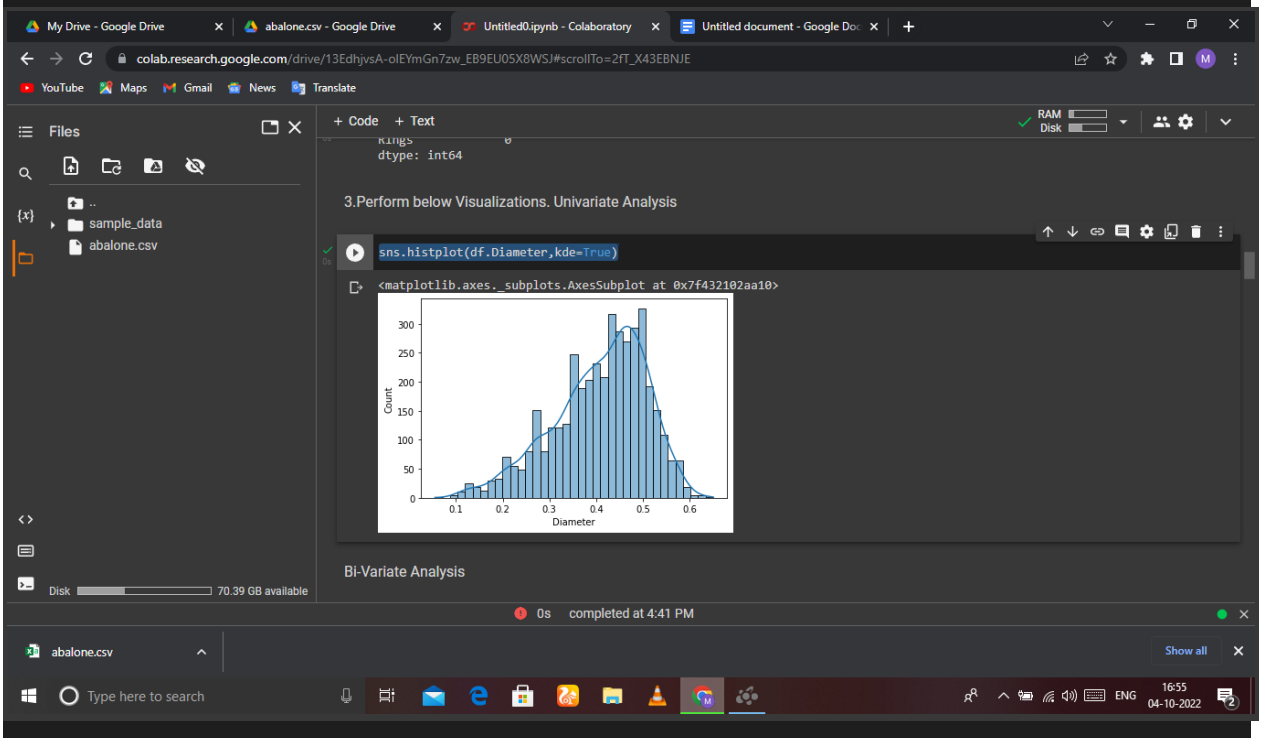

My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Do
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=0k04_PAOAcz8
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
[4177 rows x 9 columns]
Double-click (or enter) to edit
df.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
[8] df.describe()
Length Diameter Height Whole weight Shucked weight Viscera weight Shell weight Rings
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:53
04-10-2022


My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Do
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=0k04_PAOAcz8
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
[8] df.describe()
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 0.828742 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 0.760000 1.005000 29.000000
[ ] import matplotlib.pyplot as plt
import seaborn as sns
import math
[9] df.isna().sum()
Sex 0
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:53
04-10-2022



3.Perform below Visualizations. Univariate Analysis

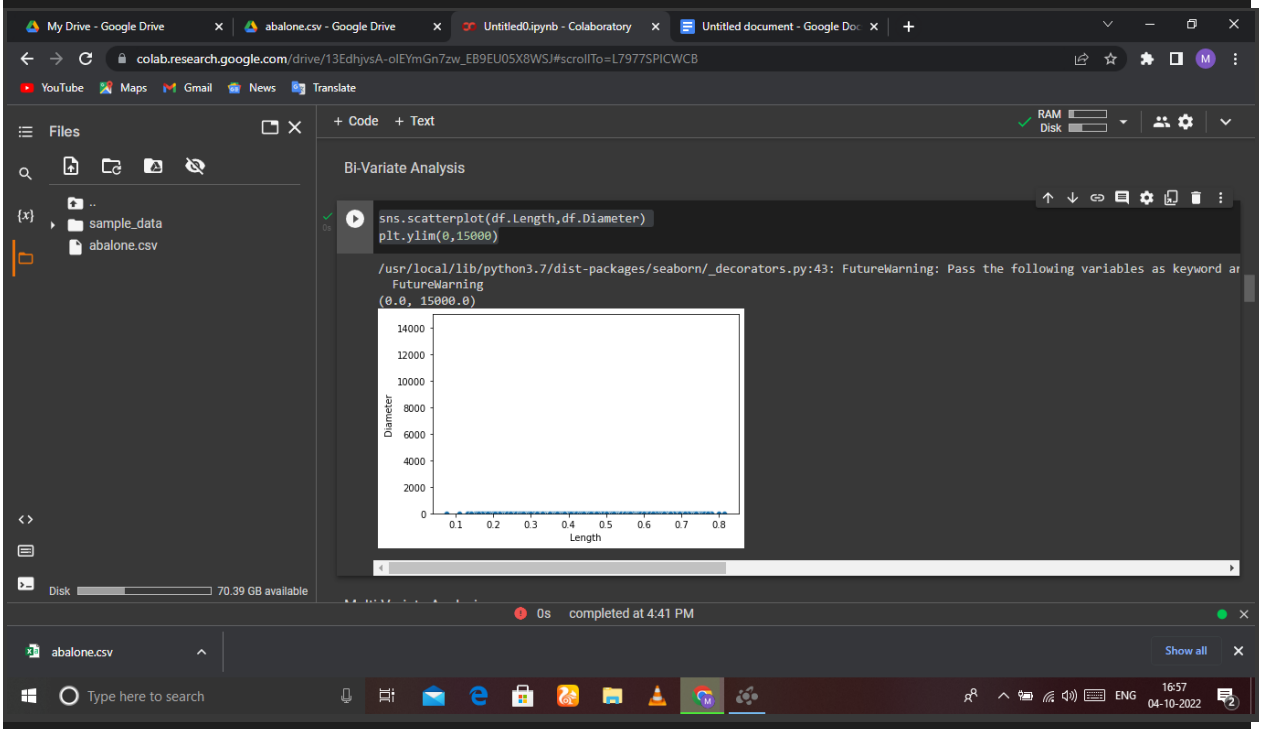
```
sns.histplot(df.Diameter,kde=True)
```



Bi-Variate Analysis

solution:

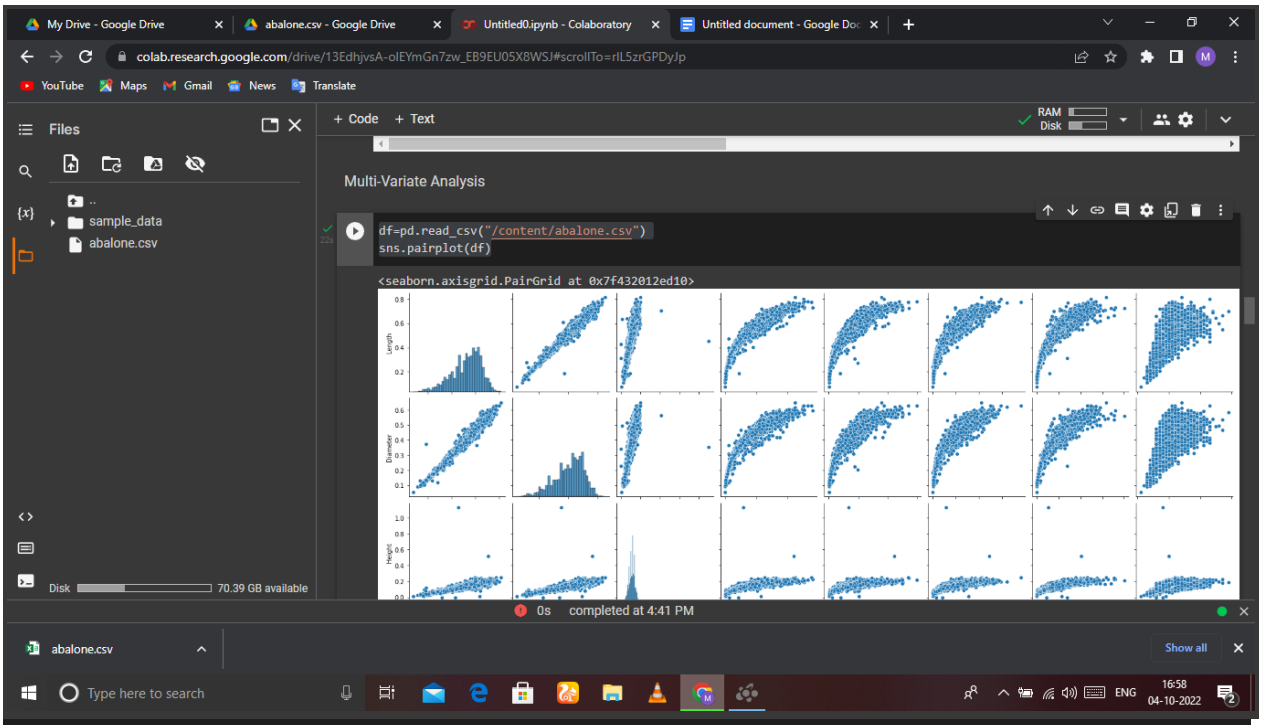
```
sns.scatterplot(df.Length,df.Diameter)
plt.ylim(0,15000)
```

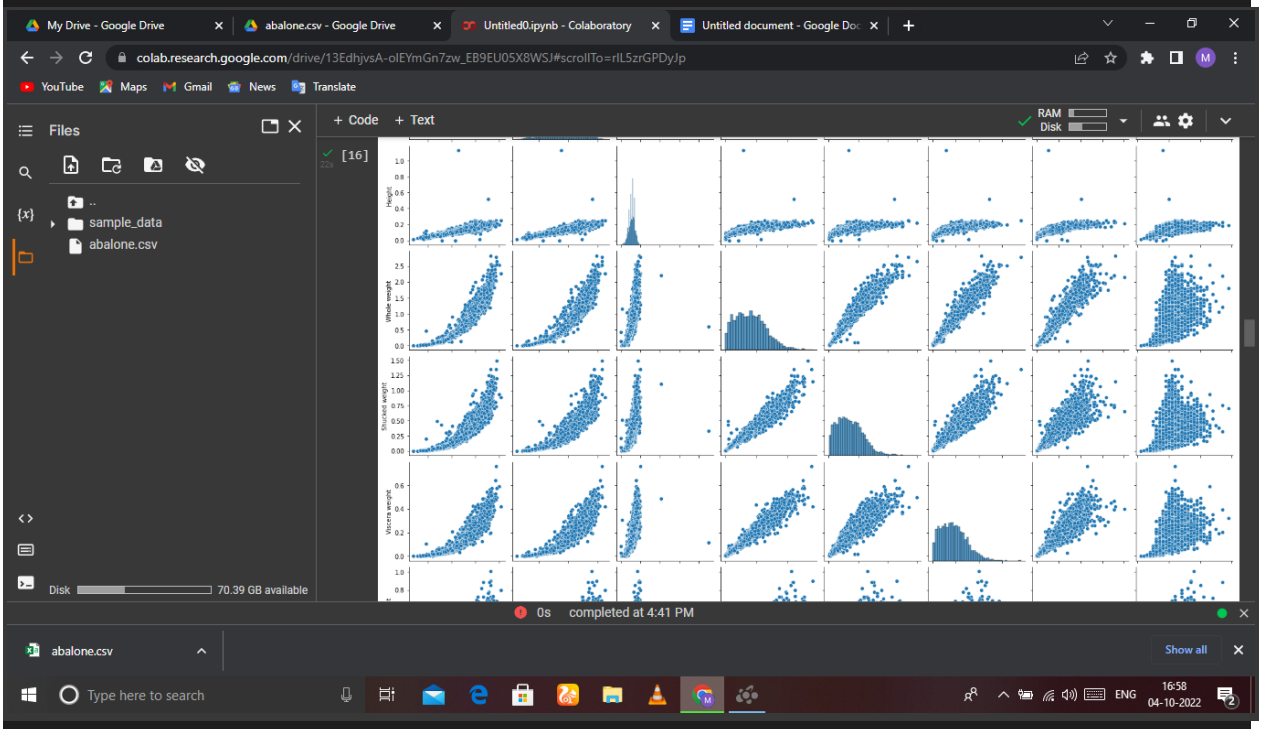


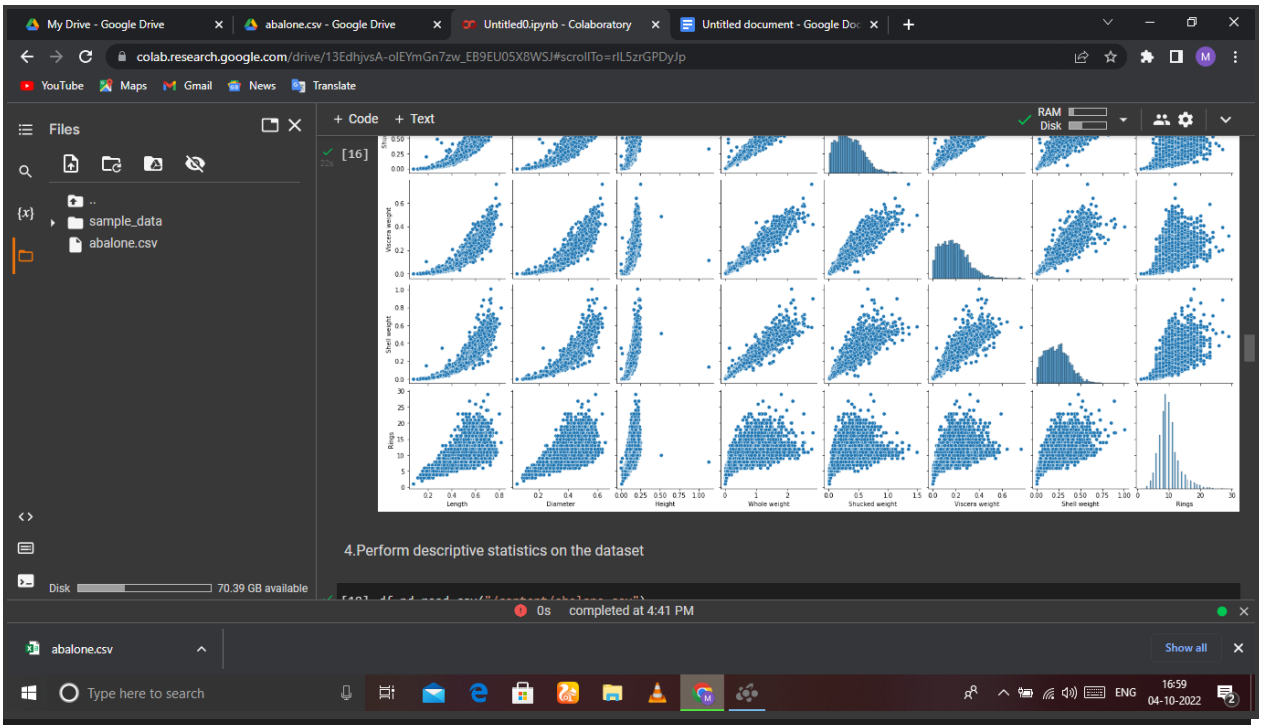
Multi-Variate Analysis

solution:

```
df=pd.read_csv("/content/abalone.csv")
sns.pairplot(df)
```


My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Doc
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=rIL5zrGPDyJp
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
Multi-Variate Analysis
df=pd.read_csv("/content/abalone.csv")
sns.pairplot(df)
<seaborn.axisgrid.PairGrid at 0x7f432012ed10>
Length
Diameter
Height
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:58
04-10-2022

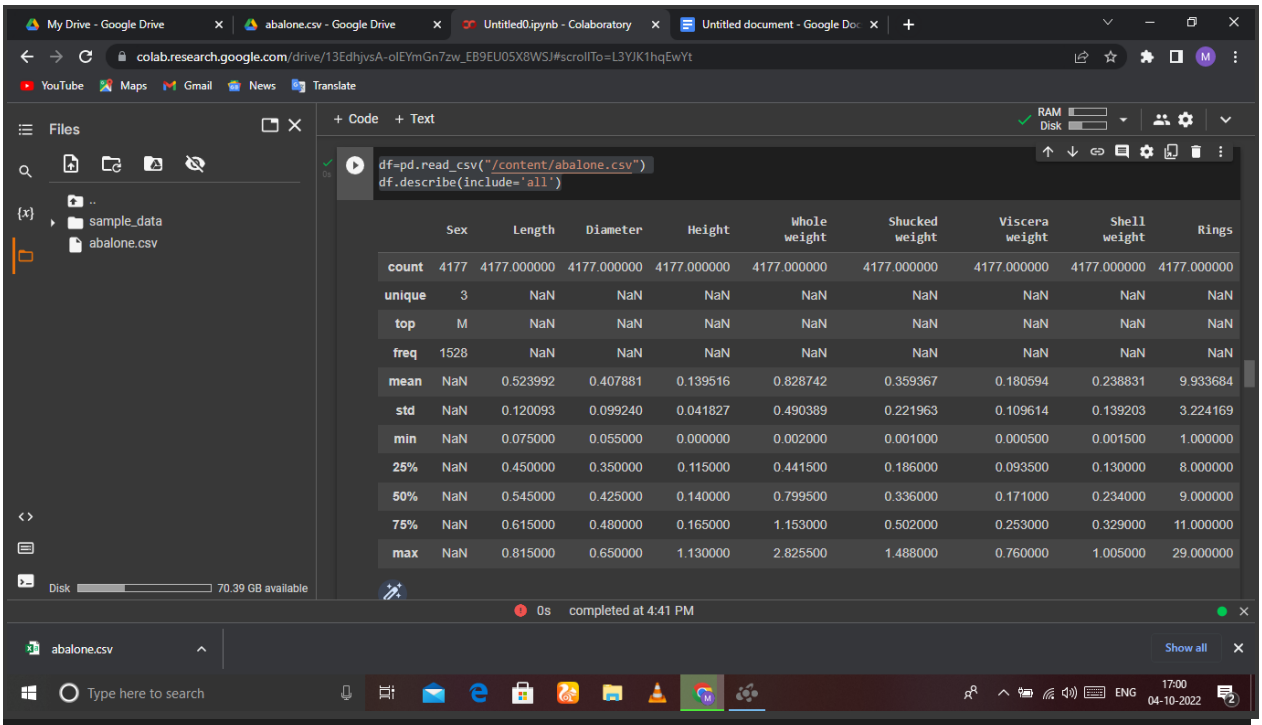

My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Doc
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=rIL5zrGPDyJp
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
[16]
Height
Whole weight
Shucked weight
Viscera weight
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
16:58
04-10-2022



4.Perform descriptive statistics on the dataset

solution:

```
df=pd.read_csv("/content/abalone.csv")
df.describe(include='all')
```

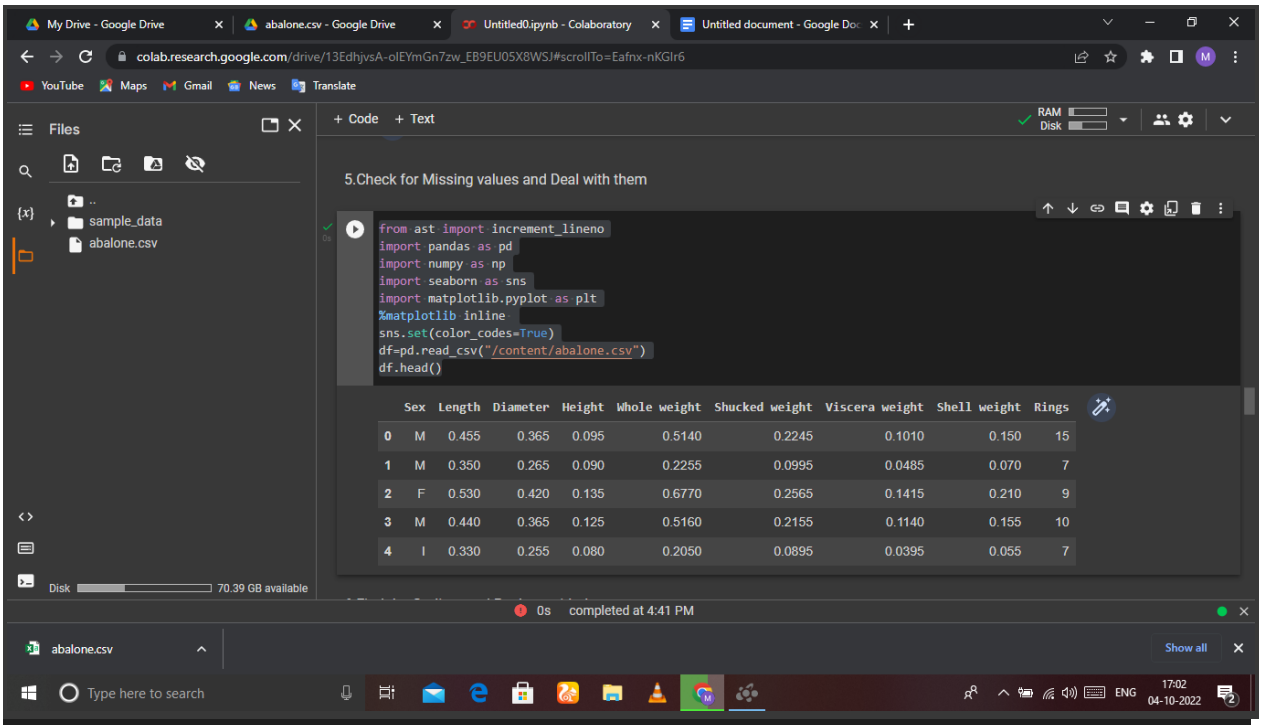


	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
count	4177	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000	4177.000000
unique	3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
top	M	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
freq	1528	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
mean	NaN	0.523992	0.407881	0.139516	0.828742	0.359367	0.180594	0.238831	9.933684
std	NaN	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0.139203	3.224169
min	NaN	0.075000	0.055000	0.000000	0.002000	0.001000	0.000500	0.001500	1.000000
25%	NaN	0.450000	0.350000	0.115000	0.441500	0.186000	0.093500	0.130000	8.000000
50%	NaN	0.545000	0.425000	0.140000	0.799500	0.336000	0.171000	0.234000	9.000000
75%	NaN	0.615000	0.480000	0.165000	1.153000	0.502000	0.253000	0.329000	11.000000
max	NaN	0.815000	0.650000	1.130000	2.825500	1.488000	0.760000	1.005000	29.000000

5.Check for Missing values and Deal with them

solution:

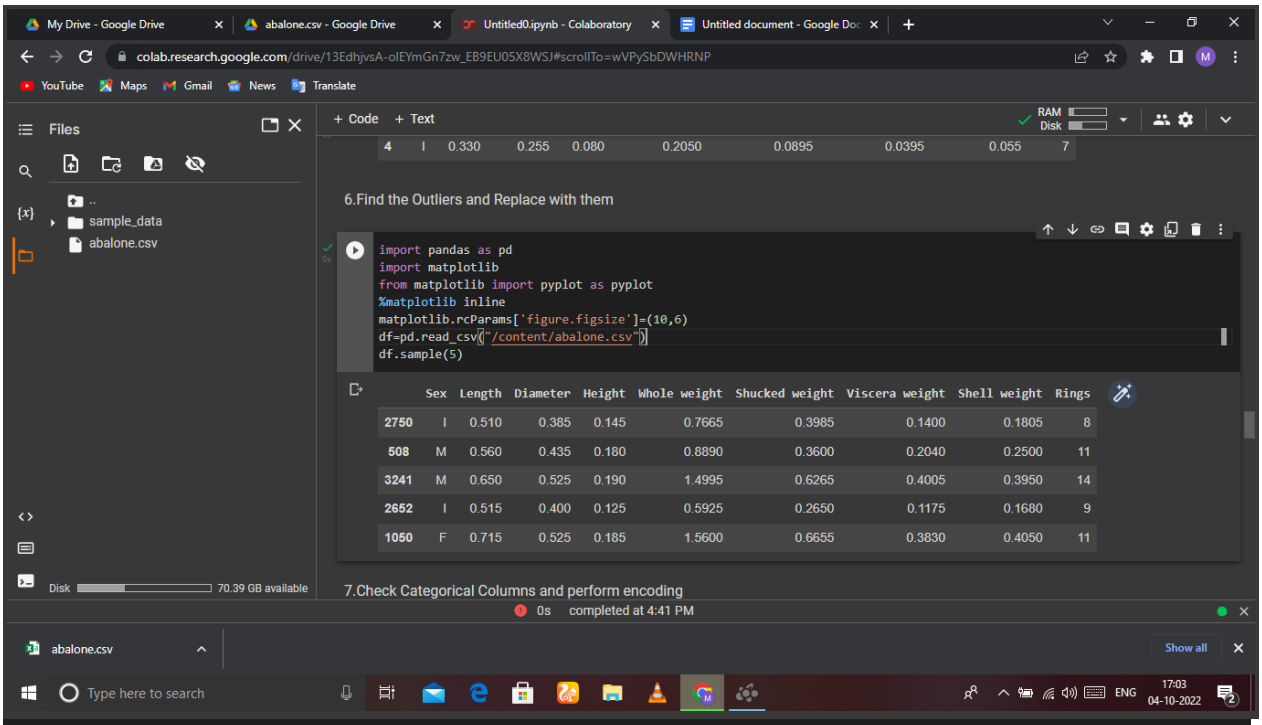
```
from ast import increment_lineno
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
df=pd.read_csv("/content/abalone.csv")
df.head()
```



6.Find the Outliers and Replace with them

solution

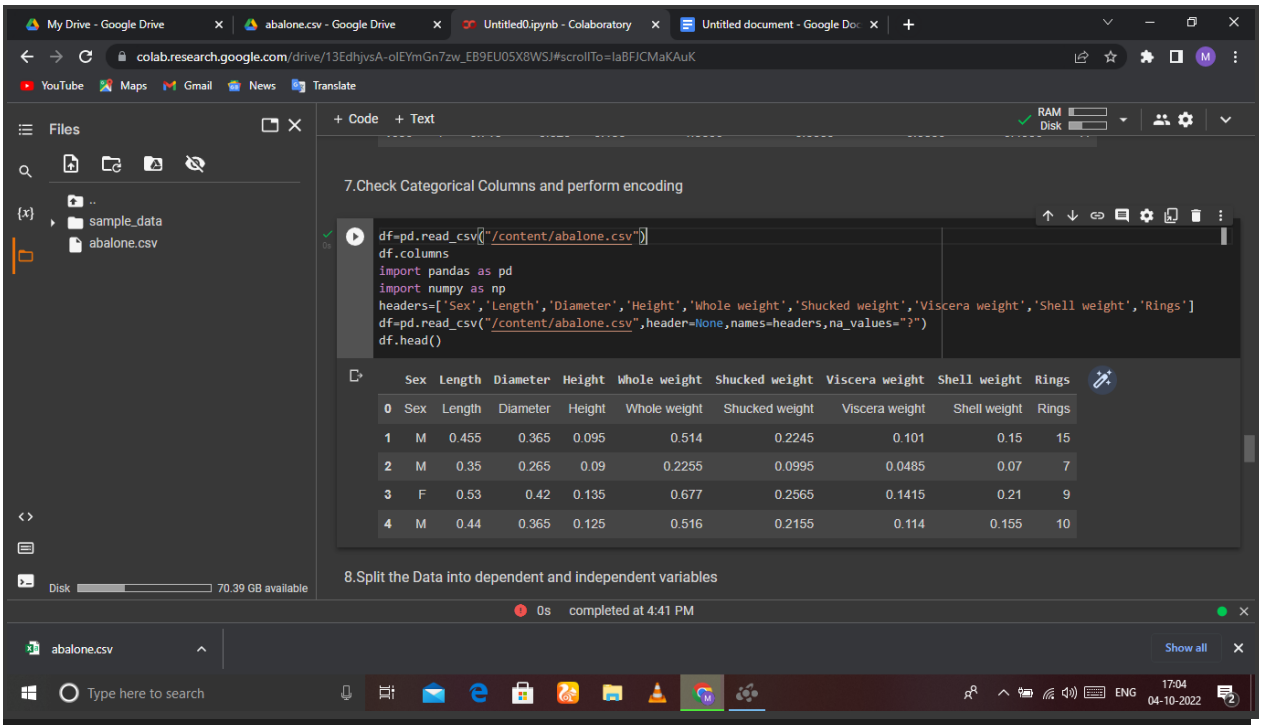
```
import pandas as pd
import matplotlib
from matplotlib import pyplot as pyplot
%matplotlib inline
matplotlib.rcParams['figure.figsize']=(10,6)
df=pd.read_csv("/content/abalone.csv")
df.sample(5)
```



7.Check Categorical Columns and perform encoding

solution:

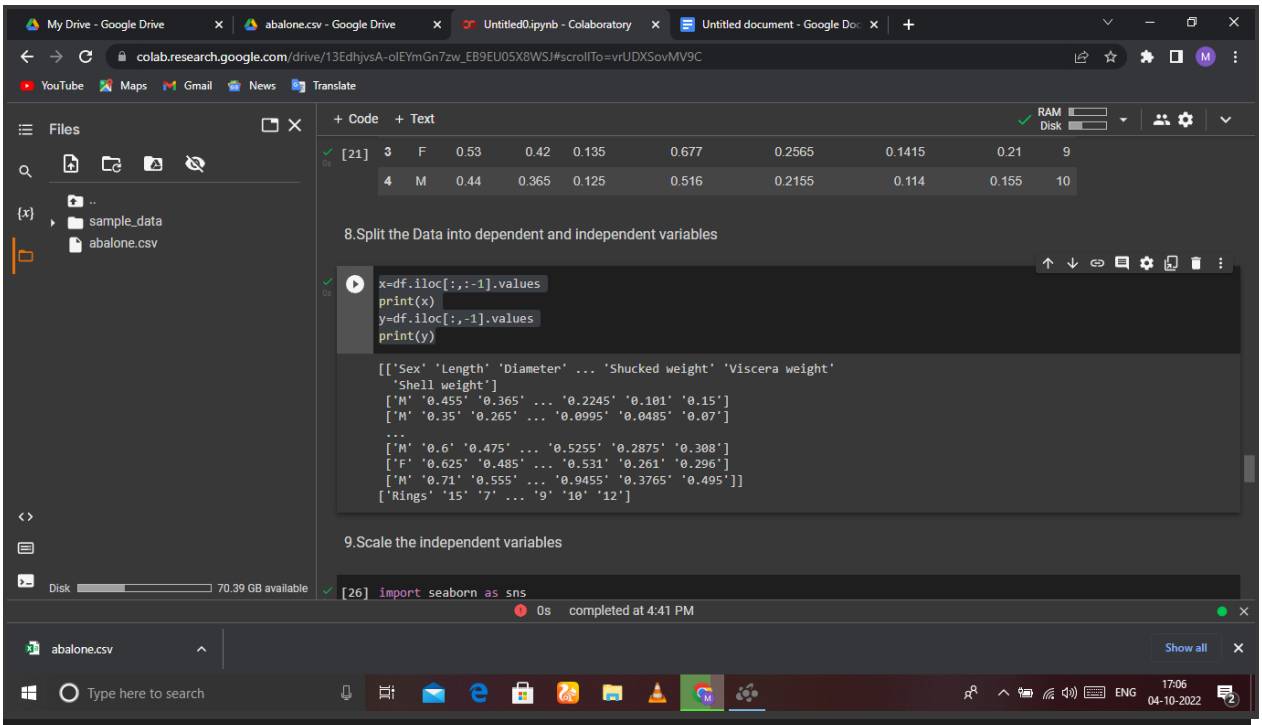
```
df=pd.read_csv("/content/abalone.csv")
df.columns
import pandas as pd
import numpy as np
headers=['Sex','Length','Diameter','Height','Whole weight','Shucked weight','Viscera weight','Shell weight','Rings']
df=pd.read_csv("/content/abalone.csv",header=None,names=headers,na_values="?")
df.head()
```



8.Split the Data into dependent and independent variables

solution:

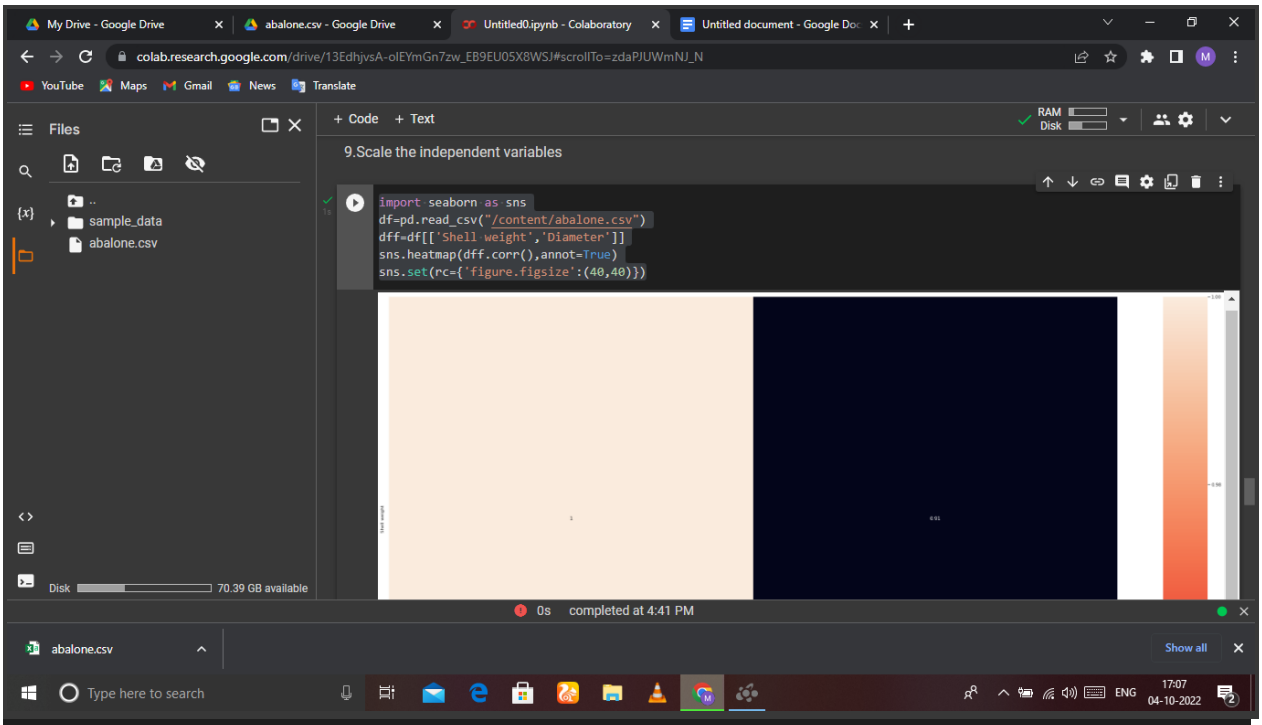
```
x=df.iloc[:,:-1].values
print(x)
y=df.iloc[:,-1].values
print(y)
```

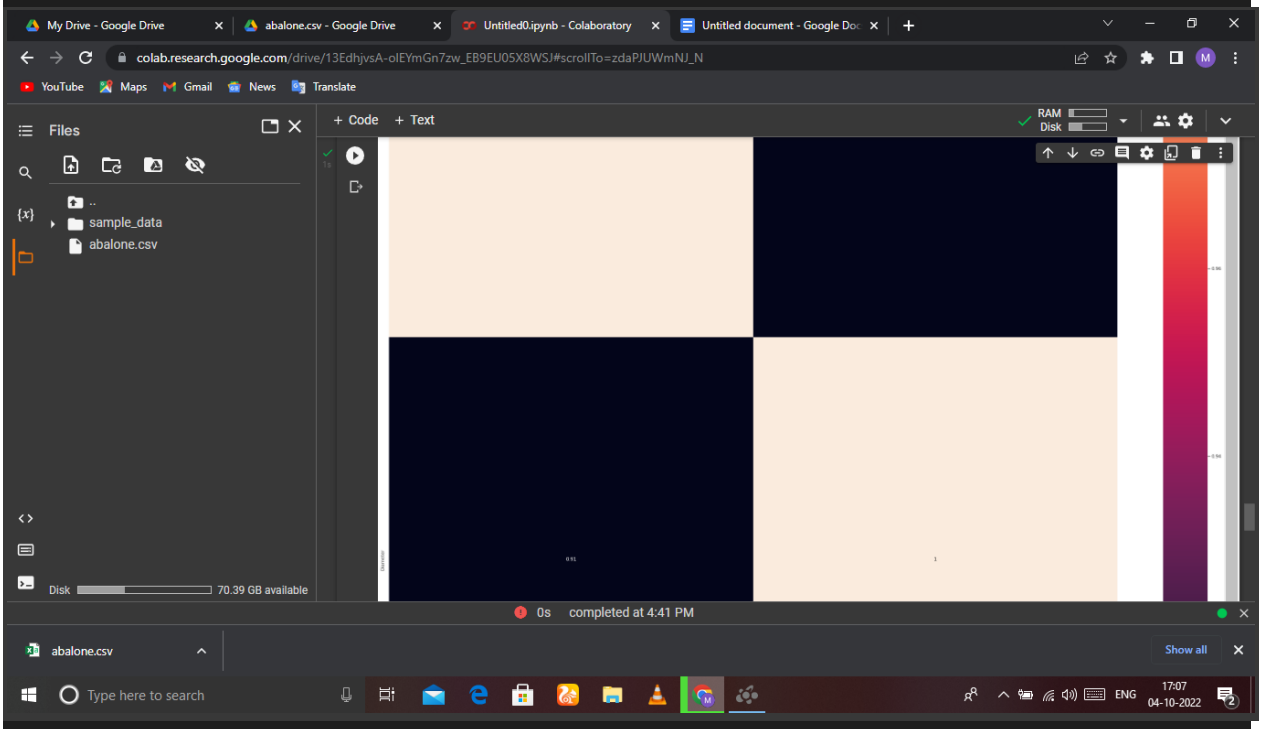


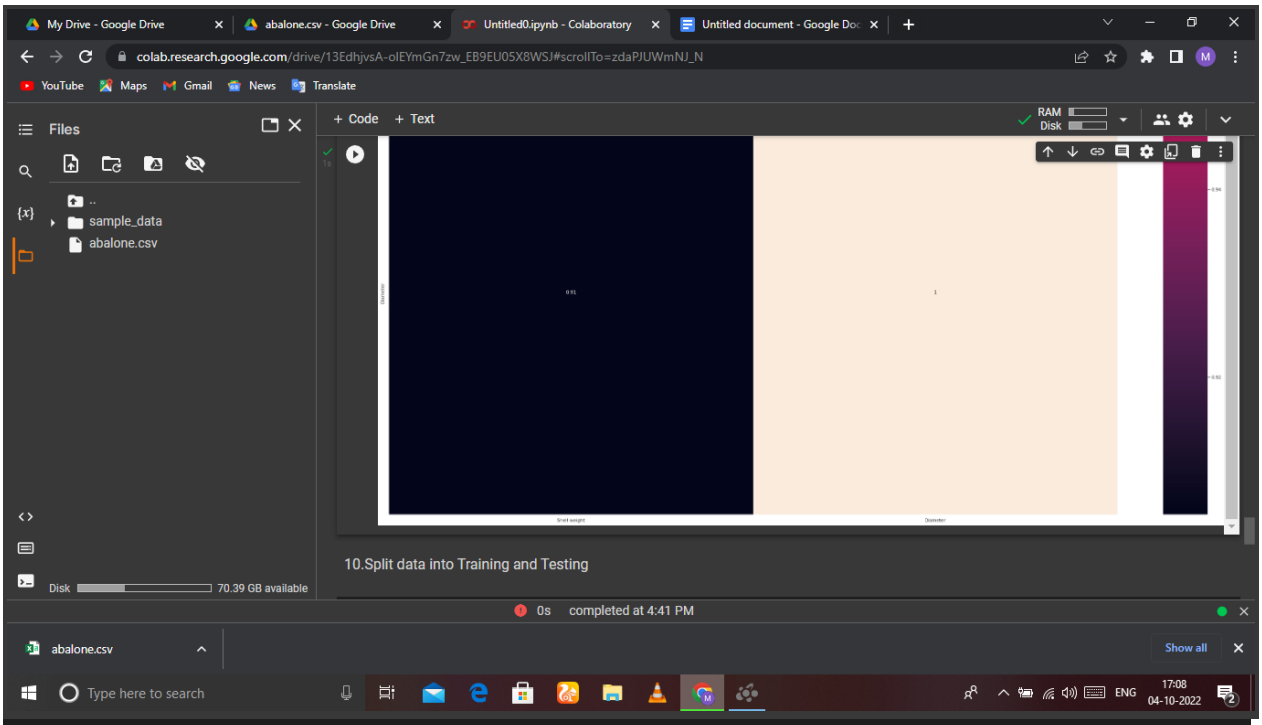
9.Scale the independent variables

solution:

```
import seaborn as sns
df=pd.read_csv("/content/abalone.csv")
dff=df[['Shell weight','Diameter']]
sns.heatmap(dff.corr(),annot=True)
sns.set(rc={'figure.figsize':(40,40)})
```


My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Doc
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=zdaPJUWmNJ_N
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
9.Scale the independent variables
import seaborn as sns
df=pd.read_csv("/content/abalone.csv")
dff=df[['Shell weight','Diameter']]
sns.heatmap(dff.corr(),annot=True)
sns.set(rc={'figure.figsize':(40,40)})
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
17:07
04-10-2022

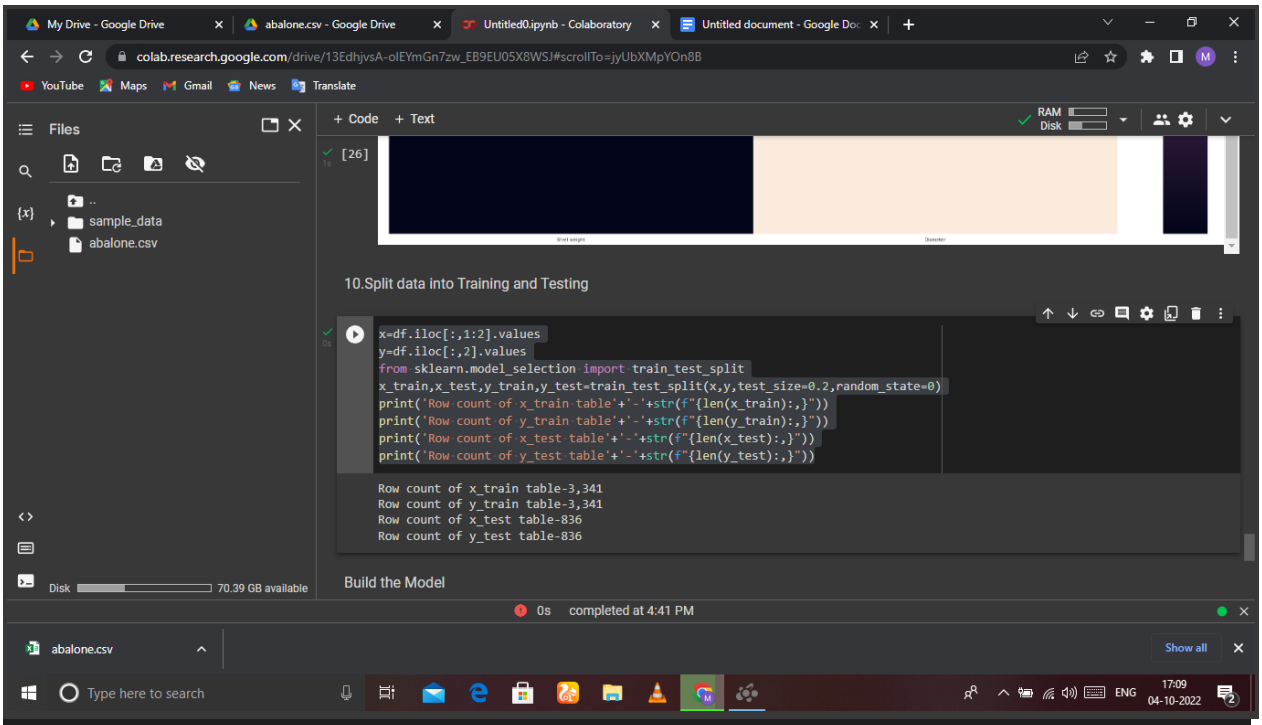

My Drive - Google Drive
abalone.csv - Google Drive
Untitled0.ipynb - Colaboratory
Untitled document - Google Doc
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ#scrollTo=zdaPJUWmNJ_N
YouTube
Maps
Gmail
News
Translate
Files
sample_data
abalone.csv
Disk 70.39 GB available
+ Code
+ Text
RAM
Disk
0s completed at 4:41 PM
abalone.csv
Show all
Type here to search
ENG
17:07
04-10-2022



10.Split data into Training and Testing

Solution:

```
x=df.iloc[:,1:2].values
y=df.iloc[:,2].values
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_st
ate=0)
print('Row count of x_train table'+'-'+str(f"{len(x_train):,}"))
print('Row count of y_train table'+'-'+str(f"{len(y_train):,}"))
print('Row count of x_test table'+'-'+str(f"{len(x_test):,}"))
print('Row count of y_test table'+'-'+str(f"{len(y_test):,}"))
```



11.Build the Model

Solution:

```
import pandas as pd
data=pd.read_csv('/content/abalone.csv')
data.head()
```

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn import svm
var=sns.heatmap(data.corr(),annot=True)
plt.show()
```

Untitled document - Google Doc | Untitled0.ipynb - Colaboratory | Untitled document - Google Doc

colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ?usp=sharing#scrollTo=yMvauW--db5c

YouTube Maps Gmail News Translate

+ Code + Text

RAM Disk

11.Build the Model

```
[20] import pandas as pd
     data=pd.read_csv('/content/abalone.csv')
     data.head()
```

	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

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn import svm
```

```
var=sns.heatmap(data.corr(),annot=True)
plt.show()
```

24s completed at 5:23 PM

Type here to search ENG 17:27 05-10-2022

Untitled document - Google Doc | Untitled0.ipynb - Colaboratory | Untitled document - Google Doc

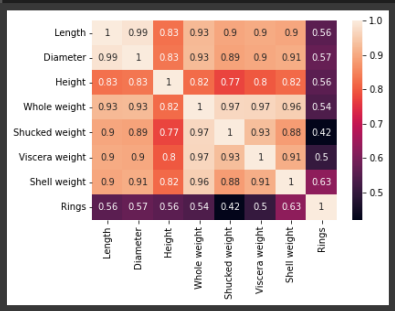
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ?usp=sharing#scrollTo=yMvauW--db5c

YouTube Maps Gmail News Translate

+ Code + Text

RAM Disk

```
var=sns.heatmap(data.corr(),annot=True)
plt.show()
```



12.Train the model

```
[21] X=data.iloc[:,:-1]
     y=data['Diameter']
```

```
[29] from sklearn.model_selection import train_test_split
```

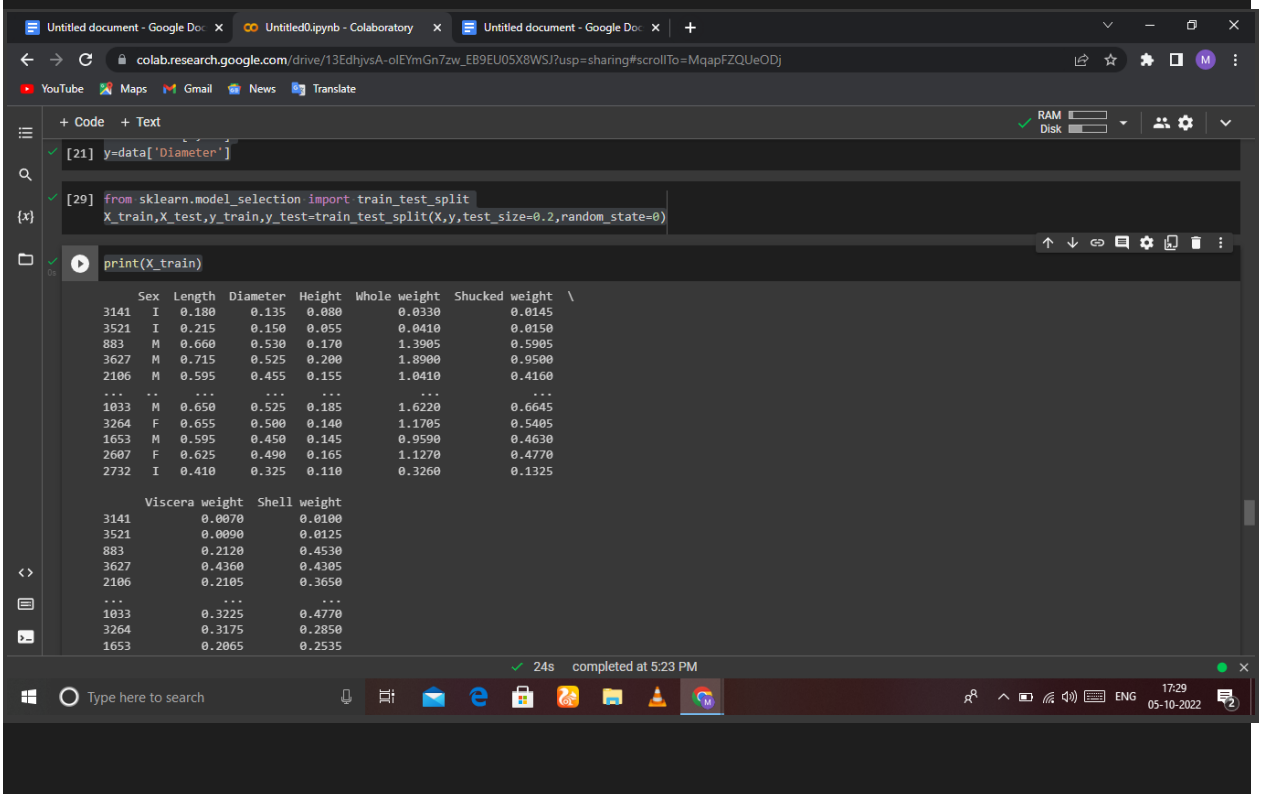
24s completed at 5:23 PM

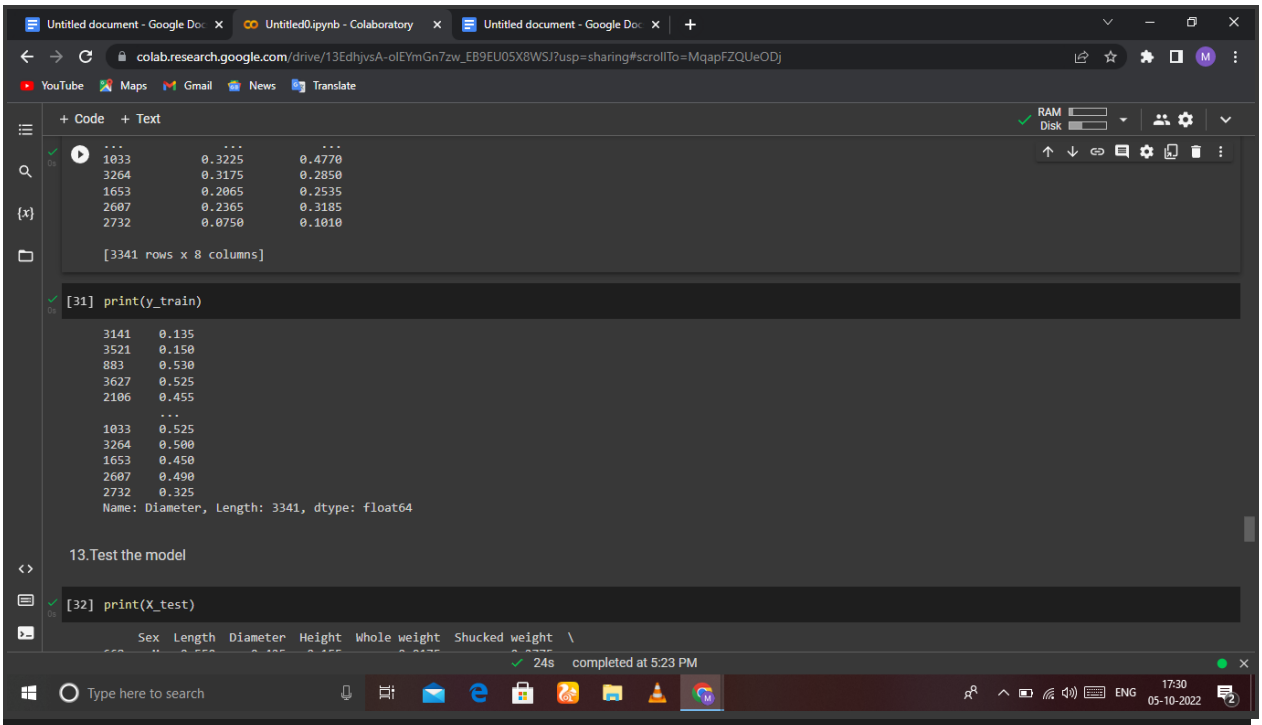
Type here to search ENG 17:28 05-10-2022

12.Train the model

solution:

```
X=data.iloc[:,:-1]
y=data['Diameter']


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_st
ate=0)
print(X_train)
print(y_train)
```



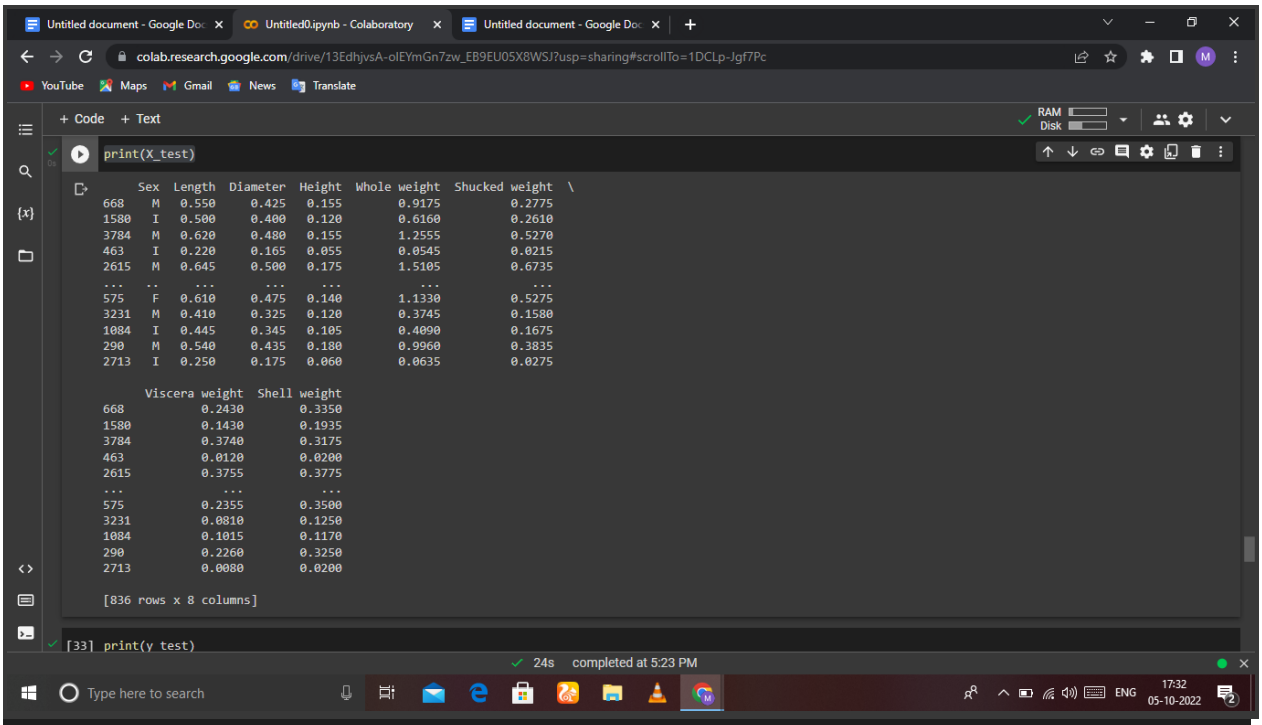


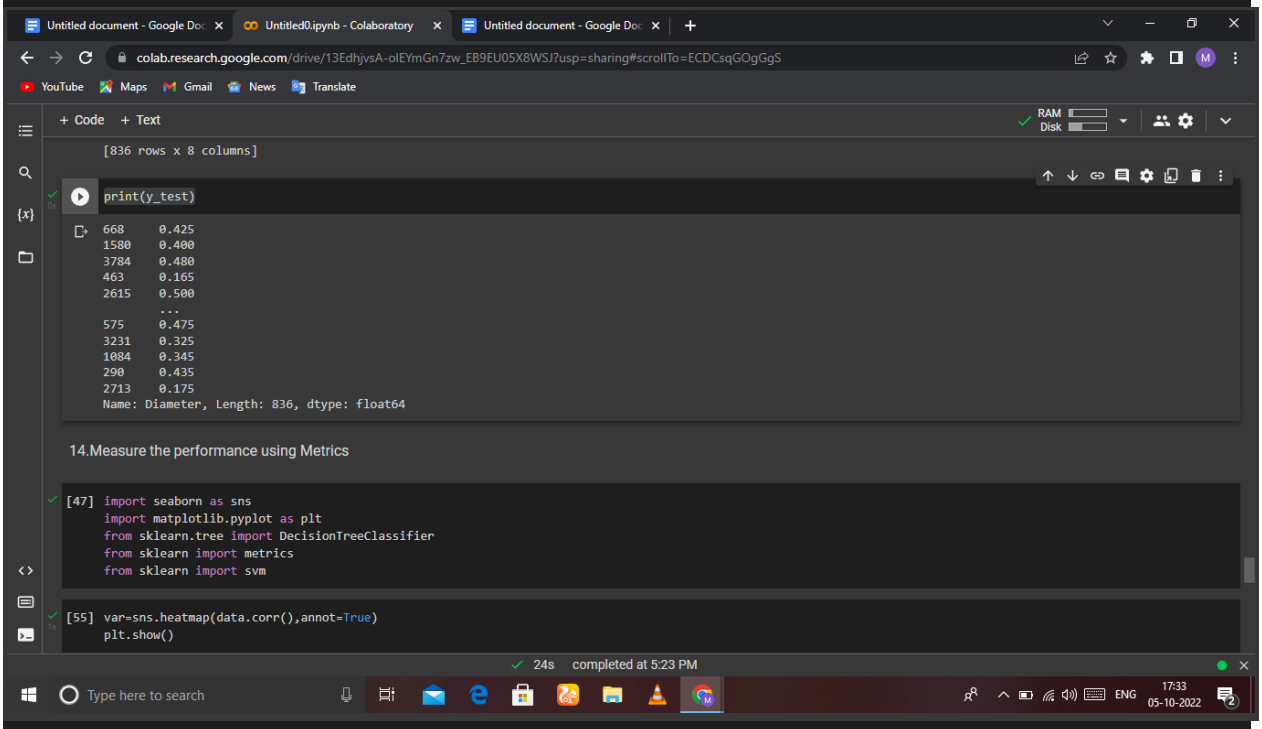
13.Test the model

Solution:

```
print(X_test)
```

```
print(y_test)
```

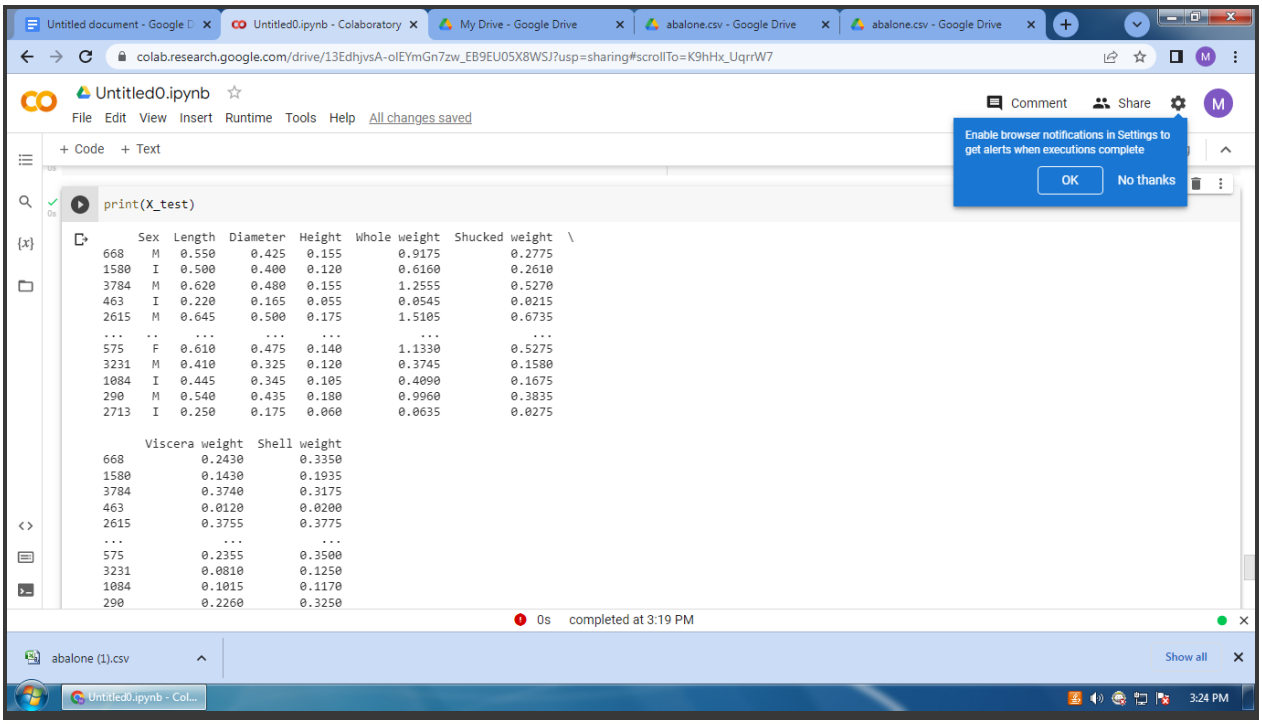


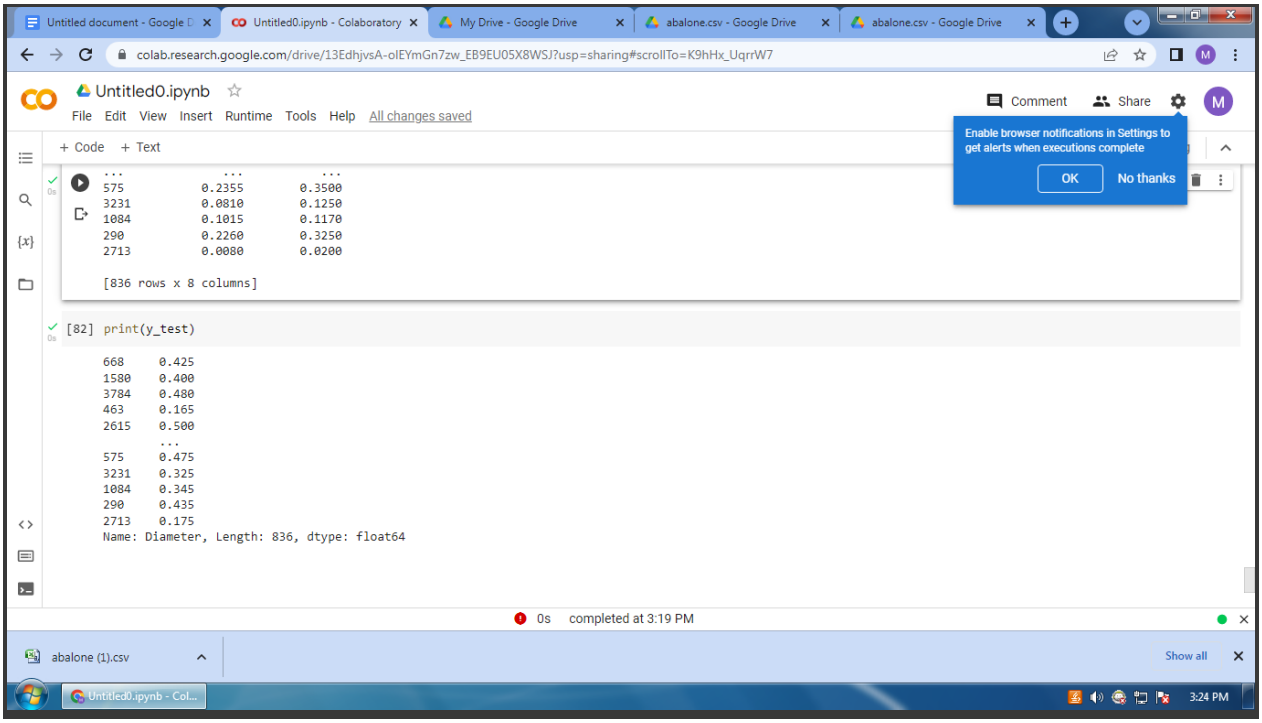


14.Measure the performance using Metrics

Solution:

```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
from sklearn import svm

model=svm.SVC(kernel='linear')

X=data.iloc[:,:-1]
y=data['Diameter']

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_st
ate=0)

print(X_test)

print(y_test)
```


Untitled document - Google D
Untitled0.ipynb - Colaboratory
My Drive - Google Drive
abalone.csv - Google Drive
abalone.csv - Google Drive
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ?usp=sharing#scrollTo=K9hHx_UqrrW7
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Enable browser notifications in Settings to get alerts when executions complete
OK
No thanks
+ Code + Text
print(X_test)
Sex Length Diameter Height Whole weight Shucked weight \
668 M 0.550 0.425 0.155 0.9175 0.2775
1580 I 0.500 0.400 0.120 0.6160 0.2610
3784 M 0.620 0.480 0.155 1.2555 0.5270
463 I 0.220 0.165 0.055 0.0545 0.0215
2615 M 0.645 0.500 0.175 1.5105 0.6735
... .
575 F 0.610 0.475 0.140 1.1330 0.5275
3231 M 0.410 0.325 0.120 0.3745 0.1580
1084 I 0.445 0.345 0.105 0.4090 0.1675
290 M 0.540 0.435 0.180 0.9960 0.3835
2713 I 0.250 0.175 0.060 0.0635 0.0275
Viscera weight Shell weight
668 0.2430 0.3350
1580 0.1430 0.1935
3784 0.3740 0.3175
463 0.0120 0.0200
2615 0.3755 0.3775
... ...
575 0.2355 0.3500
3231 0.0810 0.1250
1084 0.1015 0.1170
290 0.2260 0.3250
0s completed at 3:19 PM
abalone (1).csv
Show all
3:24 PM


Untitled document - Google D
Untitled0.ipynb - Colaboratory
My Drive - Google Drive
abalone.csv - Google Drive
abalone.csv - Google Drive
colab.research.google.com/drive/13EdhjvsA-olEYmGn7zw_EB9EU05X8WSJ?usp=sharing#scrollTo=K9hHx_UqrrW7
Untitled0.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
Enable browser notifications in Settings to get alerts when executions complete
OK
No thanks
+ Code + Text
... ...
575 0.2355 0.3500
3231 0.0810 0.1250
1084 0.1015 0.1170
290 0.2260 0.3250
2713 0.0080 0.0200
[836 rows x 8 columns]
[82] print(y_test)
668 0.425
1580 0.400
3784 0.480
463 0.165
2615 0.500
...
575 0.475
3231 0.325
1084 0.345
290 0.435
2713 0.175
Name: Diameter, Length: 836, dtype: float64
0s completed at 3:19 PM
abalone (1).csv
Show all
3:24 PM