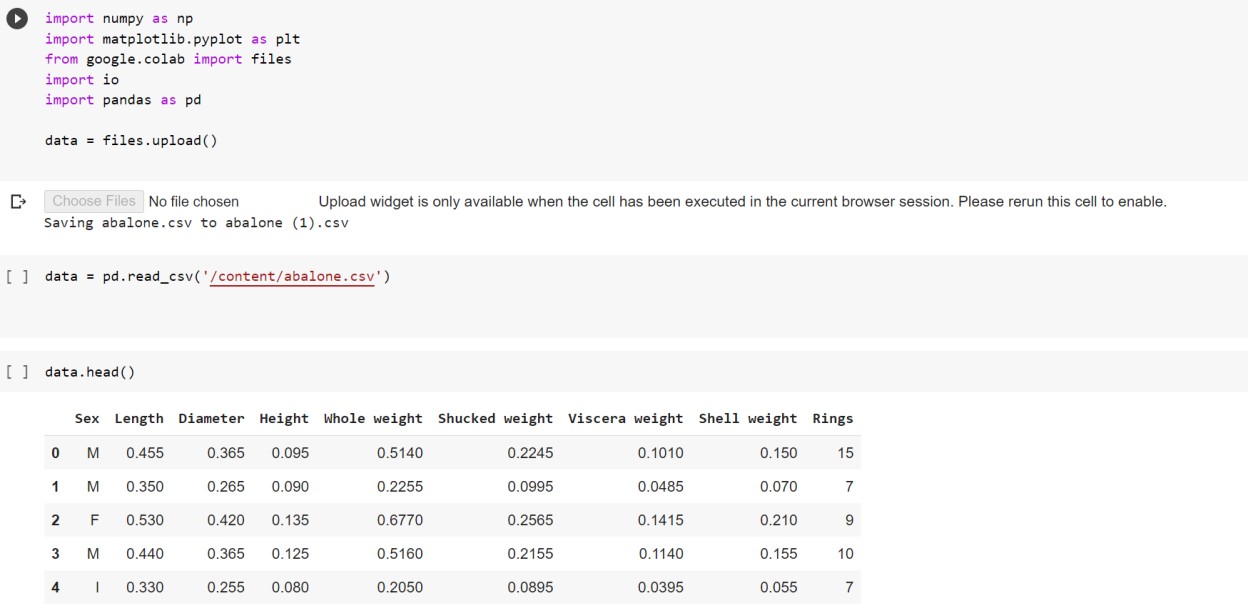
ASSIGNMENT 3

|  |  |
| --- | --- |
| Student Name | Vishnu Priya MK |
| Student Roll Number | 211519104185 |
| Maximum Marks | 10 Marks |

Question 1&2:



# Question 3:

Perform Below Visualizations.

# Univariate Analysis

* Bi-Variate Analysis

# Multivariate Analysis

Univariate Analysis

import matplotlib.pyplot as plt import seaborn as sns data.dtypes

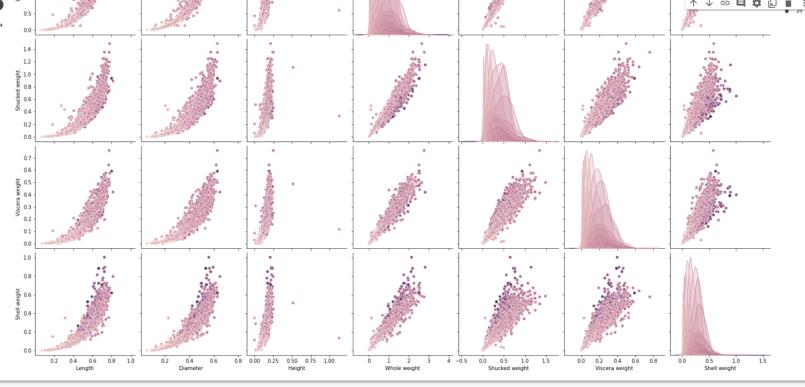
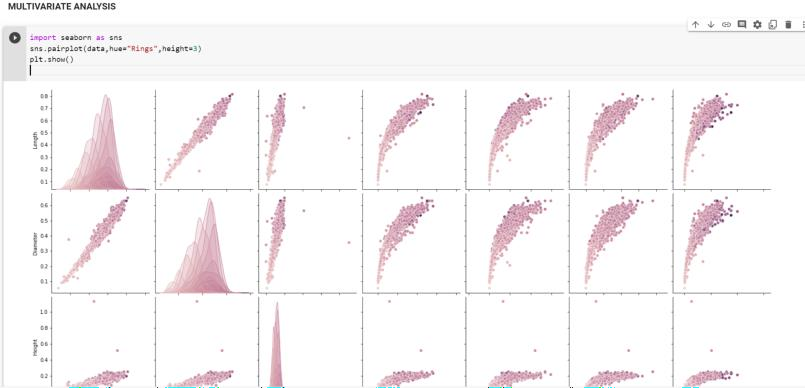
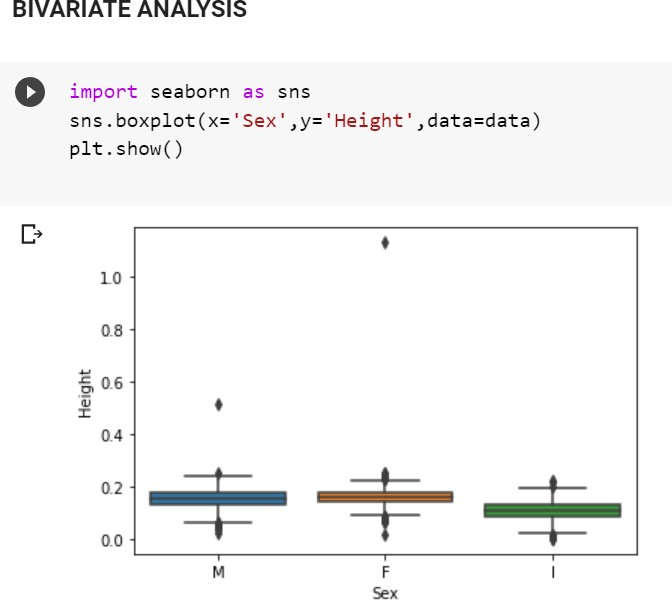
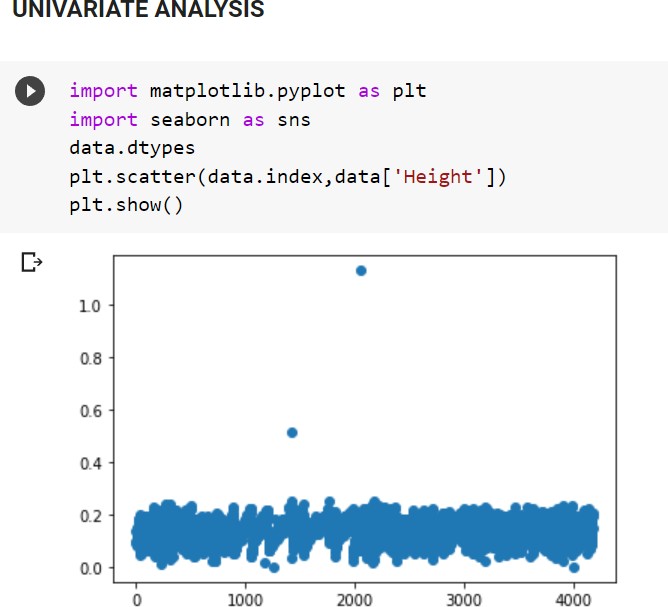
plt.scatter(data.index,data['Height']) plt.show()

# Bi-Variate Analysis

import seaborn as sns sns.boxplot(x='Sex',y='Height',data=data) plt.show()

# Multi-Variate Analysis

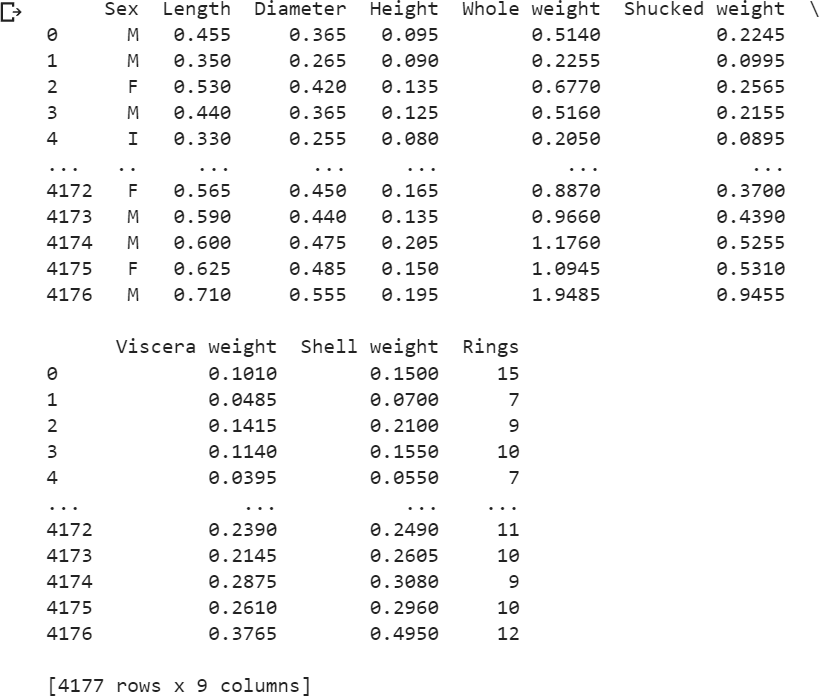
import seaborn as sns sns.pairplot(data,hue="Rings",height=3) plt.show()

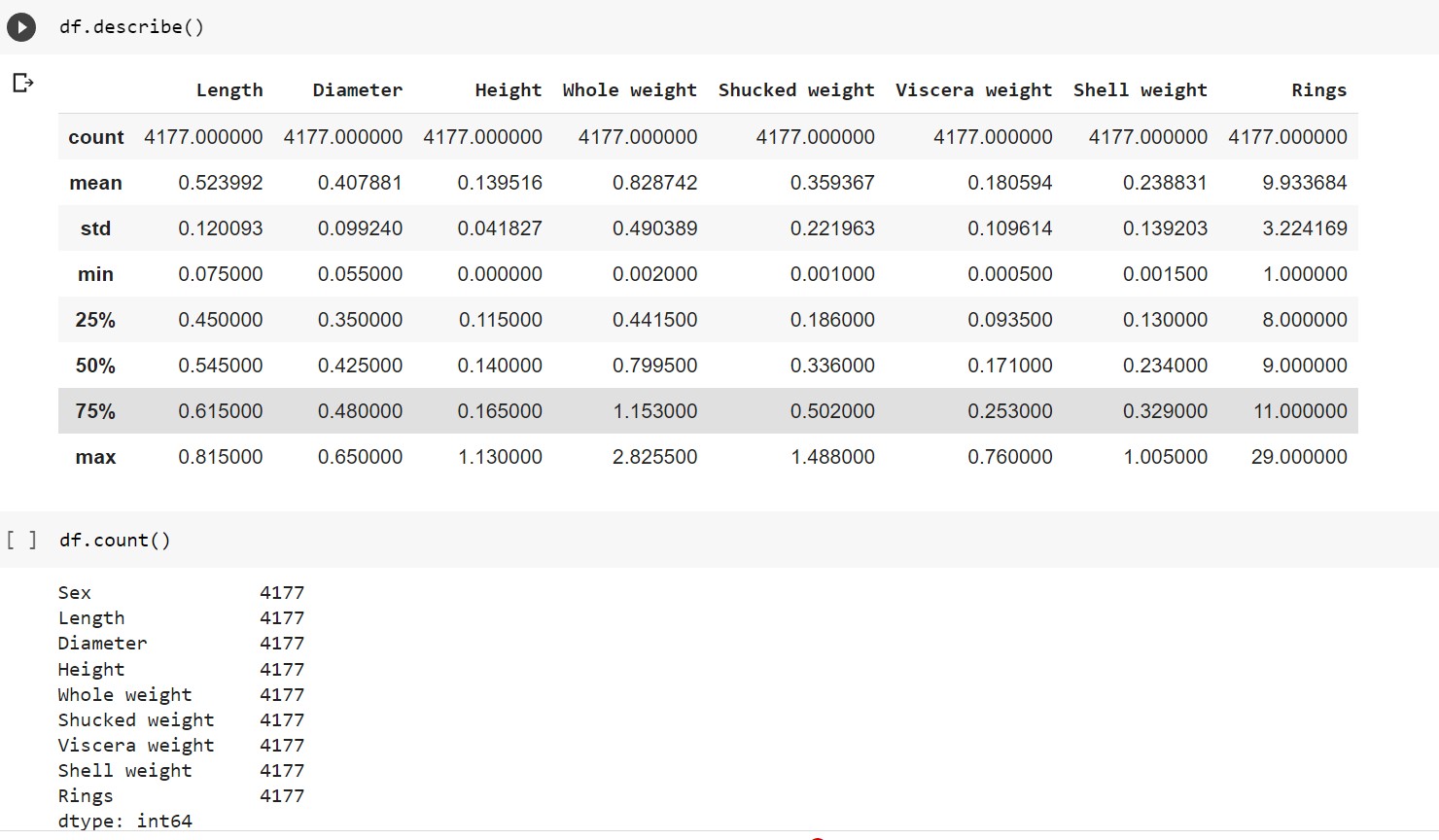


# Question 4. Perform descriptive statistics on the dataset.

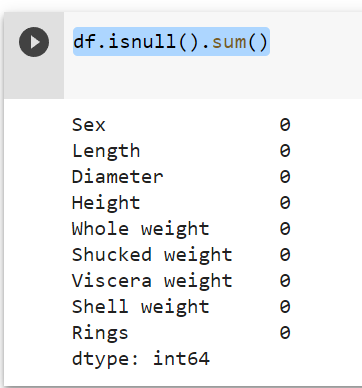
import pandas as pd

import numpy as np

df = pd.DataFrame(data) print (df) df.describe() df.count()

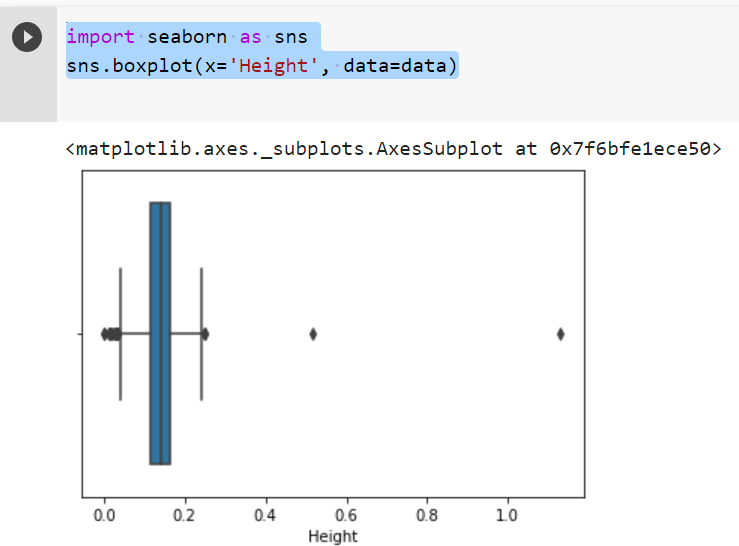


# Question 5. Check for Missing values and deal with them.

df.isnull().sum()

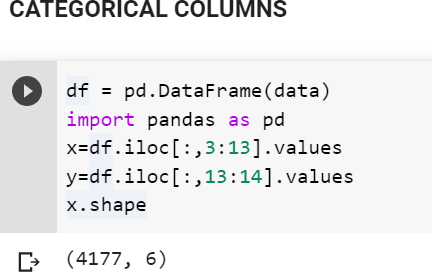
# Question 6 Find the outliers and replace them outliers

import seaborn as sns sns.boxplot(x='Height', data=data)

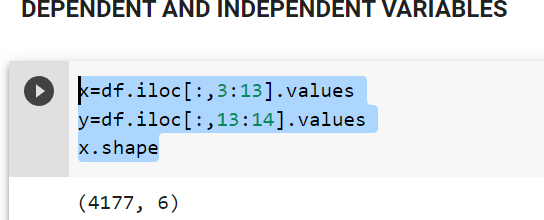


# Question 7. Check for Categorical columns and perform encoding.

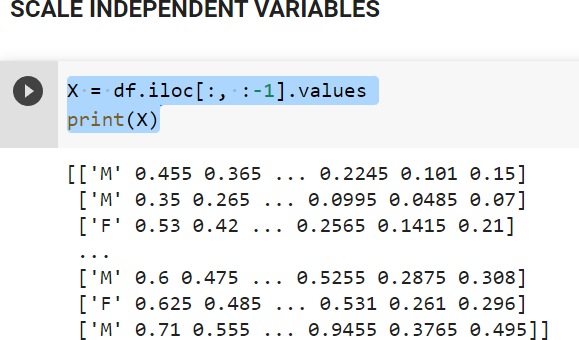
df = pd.DataFrame(data) import pandas as pd x=df.iloc[:,3:13].values y=df.iloc[:,13:14].values x.shape



# Question 8 Split the data into dependent and independent variables.

x=df.iloc[:,3:13].values y=df.iloc[:,13:14].values x.shape

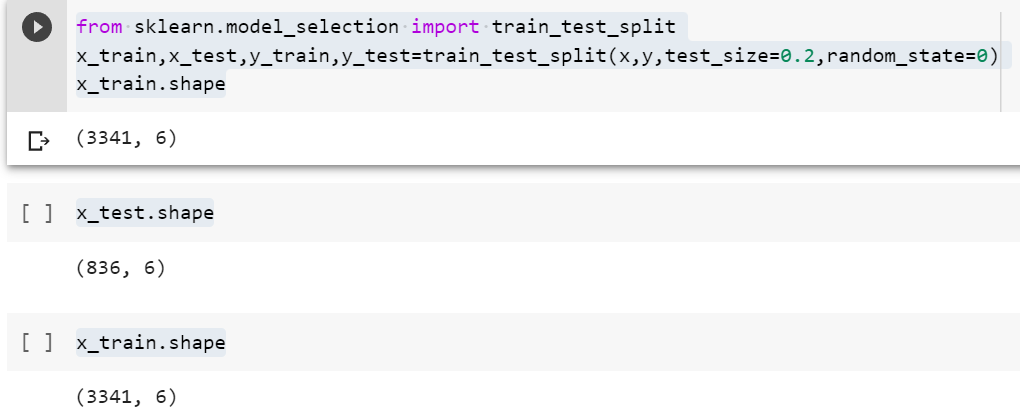
# Question 9. Scale the independent variables

X = df.iloc[:, :-1].values print(X)

# Question 10. Split the data into training and testing

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)

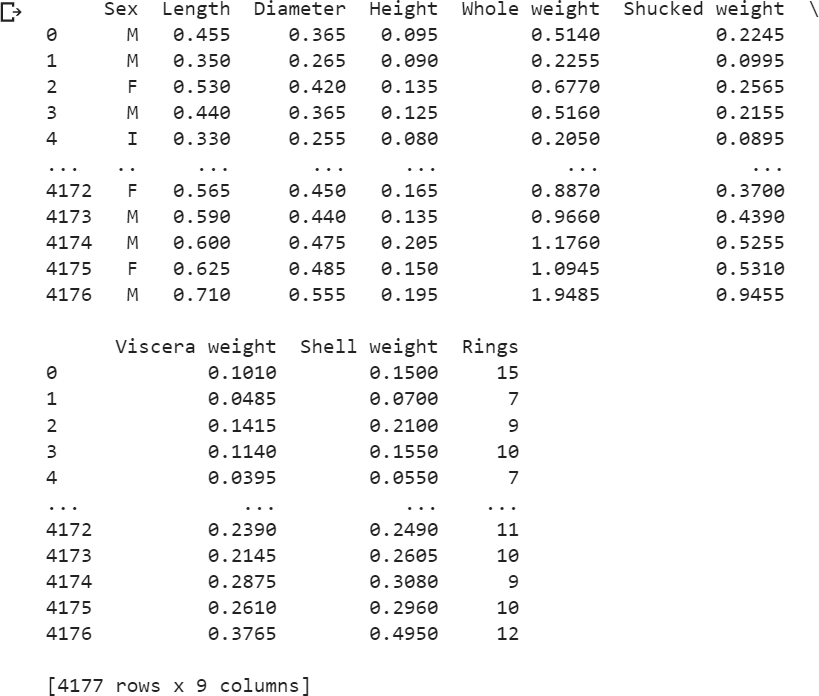
x\_train.shape x\_test.shape x\_train.shape



# Question 11. Build the Model

my\_dict=pd.read\_csv("/content/abalone.csv") df = pd.DataFrame(my\_dict)

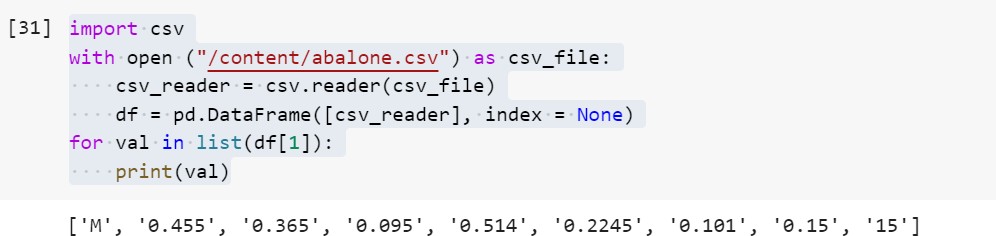
print(df)



import csv

with open ("/content/abalone.csv") as csv\_file: csv\_reader = csv.reader(csv\_file)

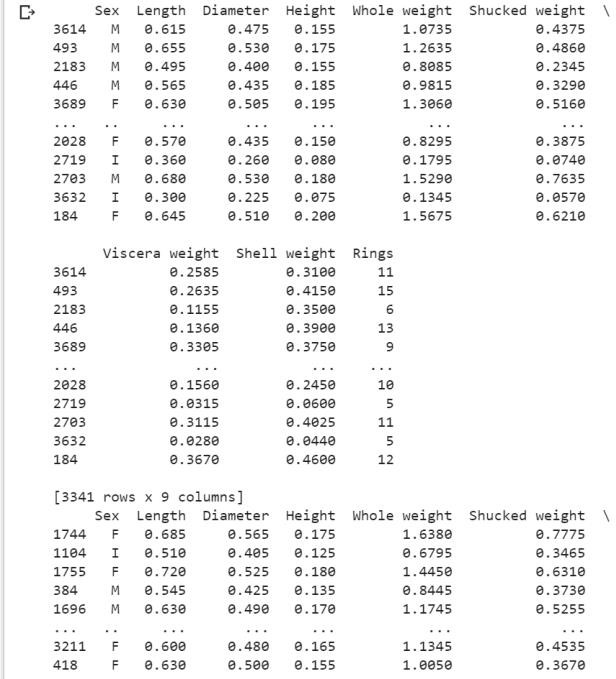
df = pd.DataFrame([csv\_reader], index = None) for val in list(df[1]):

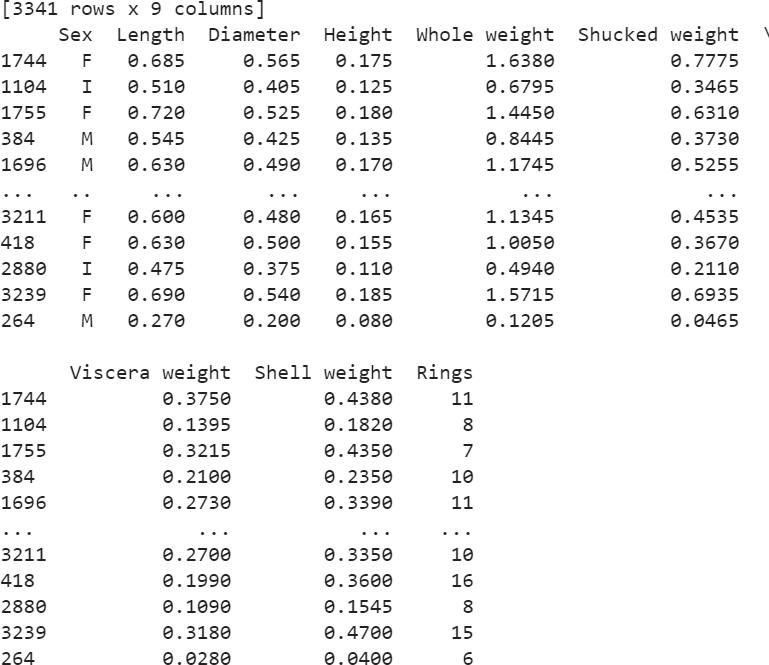
print(val)

# Question 12 & 13 Train and Test the Model

from sklearn.model\_selection import train\_test\_split

train, test = train\_test\_split(df, test\_size=0.2) print(train)

print(test)



# Question 14. Measure the performance using Metrics.

pd.crosstab(Y\_test,y\_predict) print(classification\_report(Y\_test,y\_predict))