**APPLIED DATA**

**SCIENCE**

ASSSIGNMENT-3

DONE BY.,

BALAJI KK

REG:961819106009

TEAM ID:

BATCH:B12-6A2E

from google.colab import files

uploaded=files.upload()

**abalone.csv**(text/csv) - 191962 bytes, last modified: 10/15/2022 - 100% done

Saving abalone.csv to abalone (1).csv

#Importing packages

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

data=pd.read\_csv('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 |

data=pd.read\_csv('abalone.csv')

data.shape

(4177, 9)

data=pd.read\_csv('abalone.csv')

data.tail()

SexLengthDiameterHeightWhole weightShucked weightViscera weightShell weightRings4172F0.5650.4500.1650.88700.37000.23900.2490114173M0.5900.4400.1350.96600.43900.21450.2605104174M0.6000.4750.2051.17600.52550.28750.308094175F0.6250.4850.1501.09450.53100.26100.2960104176M0.7100.5550.1951.94850.94550.37650.495012

DESCRIPTIVE STATISTICS ON THE DATASET

0s

data.describe()

| **Length** | **Diameter** | **Height** | **Whole weight** | **Shucked weight** | **Viscera weight** | **Shell weight** | **Rings** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **count** | 4177.000000 | 4177.000000 | 4177.000000 | 4177.00000[['M' 0. | 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 | 29.000 |

INDEPENDANT AND DEPENDENT VARIABLES

INDEPENDENT

data=pd.read\_csv('abalone.csv')

x=data.iloc[:,:-1].values

print(x)

[['M' 0.455 0.365 ... 0.2245 0.101 0.15]

['M' 0.35 0.265 ... 0.0995 0.0485 0.07]

['F' 0.53 0.42 ... 0.2565 0.1415 0.21]

...

['M' 0.6 0.475 ... 0.5255 0.2875 0.308]

['F' 0.625 0.485 ... 0.531 0.261 0.296]

['M' 0.71 0.555 ... 0.9455 0.3765 0.495]]

DEPENDENT

y=data.iloc[:,-1].values

print(y)

[15 7 9 ... 9 10 12]

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=0.25,random\_state=1)

print(x\_train)

print(x\_test)

print(y\_train)

print(y\_test)

[['M' 0.625 0.48 ... 0.4865 0.259 0.285]

['M' 0.48 0.38 ... 0.3 0.142 0.175]

['I' 0.2 0.145 ... 0.0125 0.0095 0.011]

...

['I' 0.32 0.24 ... 0.07 0.0265 0.0425]

['F' 0.525 0.41 ... 0.416 0.163 0.18]

['I' 0.295 0.225 ... 0.0485 0.032 0.04]]

[['F' 0.44 0.34 ... 0.188 0.087 0.13]

['M' 0.565 0.435 ... 0.5795 0.1825 0.206]

['M' 0.37 0.28 ... 0.0905 0.0585 0.075]

...

['F' 0.64 0.5 ... 0.593 0.314 0.431]

['I' 0.24 0.185 ... 0.026 0.018 0.025]

['M' 0.5 0.385 ... 0.2305 0.125 0.235]]

[10 12 4 ... 5 7 9]

[10 8 9 ... 11 6 14]

MISSING VALUES

data.isnull().sum()

Sex 0 Length 0 Diameter 0 Height 0 Whole weight 0 Shucked weight 0 Viscera weight 0 Shell weight 0 Rings 0 dtype: int64

PERFORM ENCODING

data=pd.read\_csv('abalone.csv')

from sklearn.preprocessing import LabelEncoder

le=LabelEncoder()

data["Sex"]=le.fit\_transform(data["Sex"])

data["Rings"]=le.fit\_transform(data["Rings"])

data

| **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 | 0.1500 | 14 |
| **1** | 2 | 0.350 | 0.265 | 0.090 | 0.2255 | 0.0995 | 0.0485 | 0.0700 | 6 |
| **2** | 0 | 0.530 | 0.420 | 0.135 | 0.6770 | 0.2565 | 0.1415 | 0.2100 | 8 |
| **3** | 2 | 0.440 | 0.365 | 0.125 | 0.5160 | 0.2155 | 0.1140 | 0.1550 | 9 |
| **4** | 1 | 0.330 | 0.255 | 0.080 | 0.2050 | 0.0895 | 0.0395 | 0.0550 | 6 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| **4172** | 0 | 0.565 | 0.450 | 0.165 | 0.8870 | 0.3700 | 0.2390 | 0.2490 | 10 |
| **4173** | 2 | 0.590 | 0.440 | 0.135 | 0.9660 | 0.4390 | 0.2145 | 0.2605 | 9 |
| **4174** | 2 | 0.600 | 0.475 | 0.205 | 1.1760 | 0.5255 | 0.2875 | 0.3080 | 8 |
| **4175** | 0 | 0.625 | 0.485 | 0.150 | 1.0945 | 0.5310 | 0.2610 | 0.2960 | 9 |
| **4176** | 2 | 0.710 | 0.555 | 0.195 | 1.9485 | 0.9455 | 0.3765 | 0.4950 | 11 |

4177 rows × 9 columns

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 int64

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(2)

memory usage: 293.8 KB

import seaborn as sns

import matplotlib.pyplot as plt

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

plt.show()

