Assignment -4

Assignment Date	27 October 2022
Student Name	Sangeetha E
Student Roll Number	812419104709
Maximum Marks	2 Marks

+ Code   + Text

Double-click (or enter) to edit

▾ IMPORTING LIBRARIES

[ ]

Double-click (or enter) to edit

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

```
[ ] from google.colab import drive
    drive.mount('/content/drive')

    Mounted at /content/drive
```

```
[ ] #load the data
    df = pd.read_csv('/content/drive/My Drive/Machine Learning/abalone.csv')
```

```
[ ] df
```

+ Code   + Text

```
  #load the data
  df = pd.read_csv('/content/drive/My Drive/Machine Learning/abalone.csv')
```

```
[ ] df
```

	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	0.1500	15
1	M	0.350	0.265	0.090	0.2255	0.0995	0.0485	0.0700	7
2	F	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.2100	9
3	M	0.440	0.365	0.125	0.5160	0.2155	0.1140	0.1550	10
4	I	0.330	0.255	0.080	0.2050	0.0895	0.0395	0.0550	7
...	...	...	...	...	...	...	...	...	...
4172	F	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	11
4173	M	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	10
4174	M	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	9
4175	F	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	10
4176	M	0.710	0.555	0.195	1.9485	0.9455	0.3765	0.4950	12

4177 rows × 9 columns

```
[ ] df.head()
```
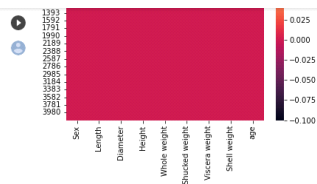
| max | 0.815000 | 0.650000 | 1.130000 | 2.825500 | 1.488000 | 0.760000 | 1.005000 | 29.000000 |

```python
df['age'] = df['Rings']+1.5
df = df.drop('Rings', axis = 1)
```

```python
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.feature_selection import SelectKBest
from sklearn.metrics import r2_score, mean_squared_error
import warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
```

UNIVARIATE ANALYSIS

```python
sns.heatmap(df.isnull())
```

```
<AxesSubplot:>
```
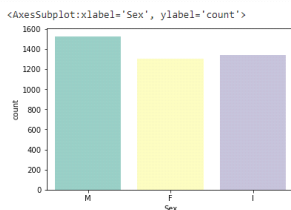


```python
df.info()
```

```
<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   age            4177 non-null   float64
dtypes: float64(8), object(1)
memory usage: 293.8+ KB
```

```
 6   Viscera weight 4177 non-null   float64
 7   Shell weight   4177 non-null   float64
 8   age            4177 non-null   float64
dtypes: float64(8), object(1)
memory usage: 293.8+ KB
```
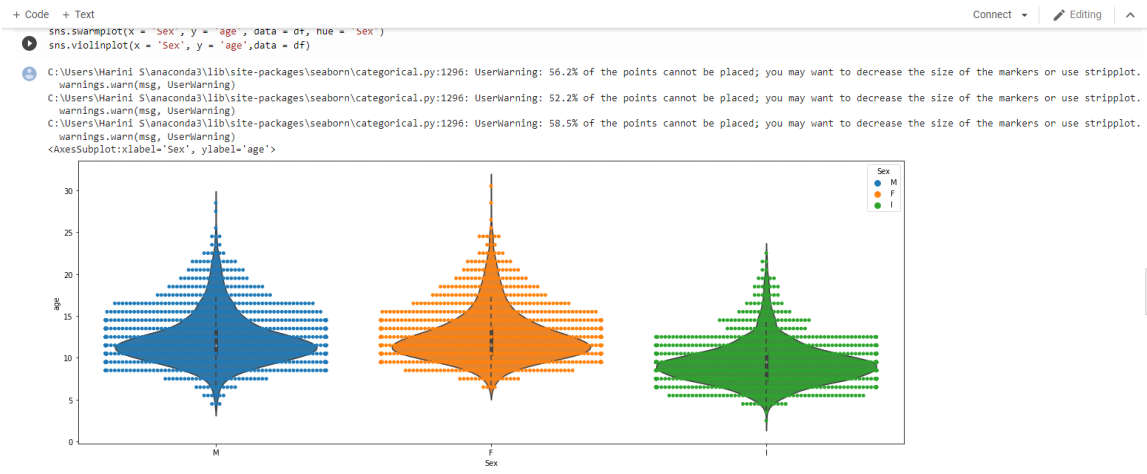
```python
sns.countplot(x = 'Sex', data = df, palette = 'Set3')
```

```
<AxesSubplot:xlabel='Sex', ylabel='count'>
```



```python
plt.figure(figsize = (20,7))
sns.swarmplot(x = 'Sex', y = 'age', data = df, hue = 'Sex')
sns.violinplot(x = 'Sex', y = 'age',data = df)
```

```
C:\Users\Harini S\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 56.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\Harini S\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 52.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
```

```
sns.swarmplot(x = 'Sex', y = 'age', data = df, hue = 'Sex')
sns.violinplot(x = 'Sex', y = 'age',data = df)
```
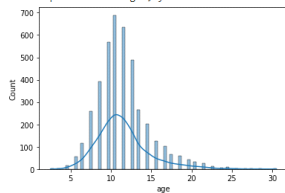
```
C:\Users\Harini S\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 56.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\Harini S\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 52.2% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
C:\Users\Harini S\anaconda3\lib\site-packages\seaborn\categorical.py:1296: UserWarning: 58.5% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.
  warnings.warn(msg, UserWarning)
<AxesSubplot:xlabel='Sex', ylabel='age'>
```



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

```
<AxesSubplot:xlabel='age', ylabel='Count'>
```


▸ BIVARIATE ANALYSIS

[ ] ↳ 3 cells hidden

▸ MULTI VARIATE ANALYSIS

[ ] ↳ 1 cell hidden

▸ BIVARIATE ANALYSIS

↳ 3 cells hidden

▸ MULTI VARIATE ANALYSIS

[ ] ↳ 1 cell hidden

▸ MISSING VALUES

[ ] ↳ 3 cells hidden

▸ OUTLIERS

[ ] ↳ 18 cells hidden

▸ CATEGORICAL COLUMNS

SCALING THE INDEPENDENT VARIABLE

▶ *2 cells hidden*

SPLITING THE DATA

[ ] *2 cells hidden*

BUILDING MODEL

[ ] *4 cells hidden*

TRAINING THE MODEL

[ ] *3 cells hidden*

TESTING THE MODEL

[ ] *5 cells hidden*