

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib as mp
from scipy import stats
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
```

```
In [13]: ds=pd.read_csv('table1_3.csv')
```

```
In [14]: ds.head()
```

```
Out[14]:
```

	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	0.1050	0.100	15
1	M	0.360	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	0.2195	0.1140	0.155	10
4	I	0.320	0.255	0.080	0.2260	0.0995	0.0295	0.095	7

```
In [15]: ds.tail()
```

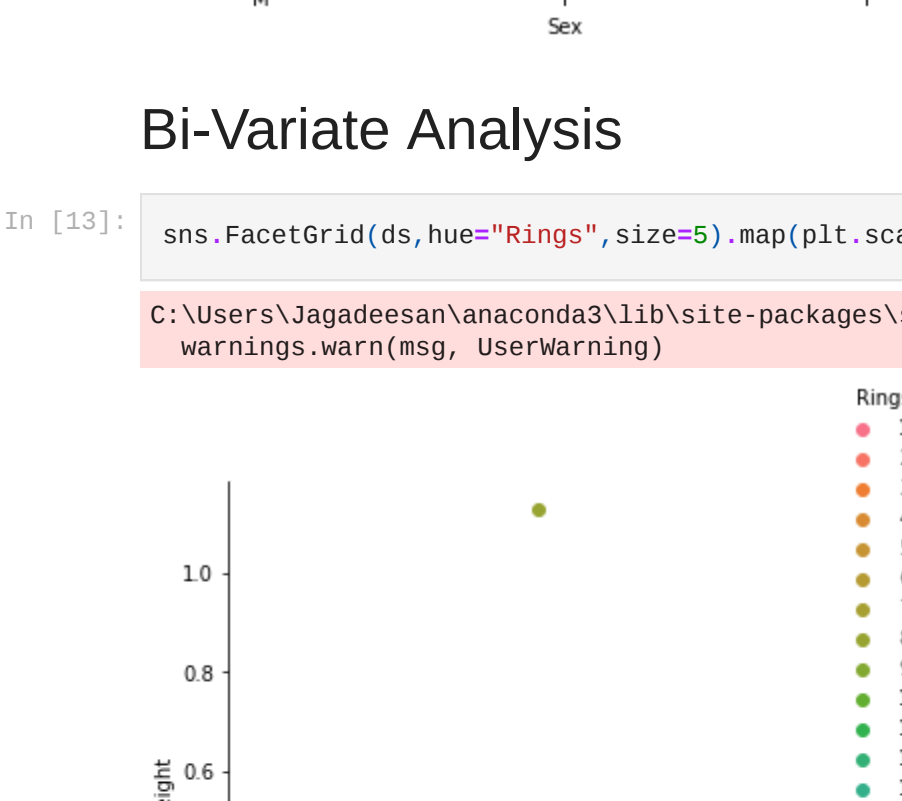
```
Out[15]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
4172	F	0.565	0.460	0.165	0.8870	0.3760	0.2360	0.340	11
4173	F	0.560	0.440	0.155	0.9050	0.4080	0.2460	0.290	10
4174	M	0.800	0.475	0.205	1.1760	0.5355	0.3875	0.300	9
4175	F	0.625	0.485	0.150	1.0845	0.5310	0.2610	0.290	10
4176	M	0.710	0.555	0.195	1.9465	0.9455	0.3765	0.450	12

Univariate Analysis

```
In [7]: df.info()
df.describe()
df.groupby('Sex').info()
df.groupby('Sex').describe()
```

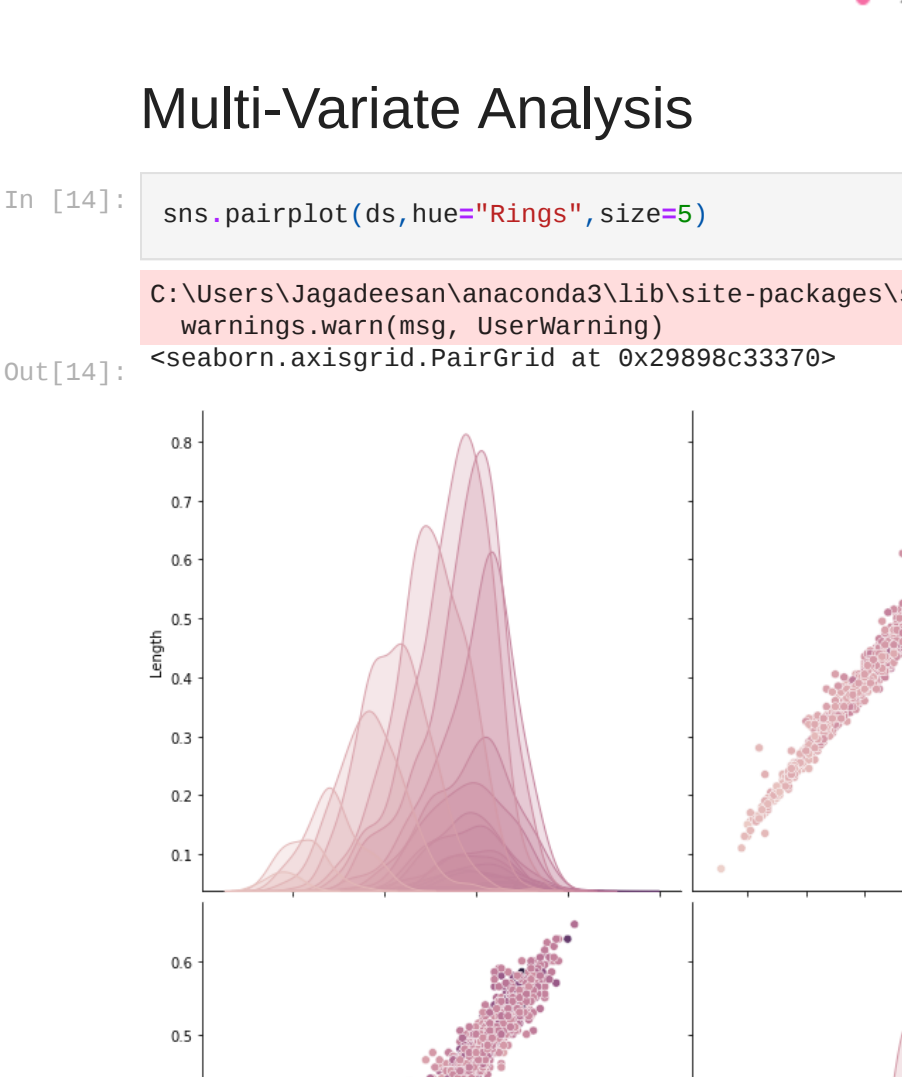
```
In [11]: plt.plot(df['Sex'],np.zeros_like(df['Sex']))
plt.plot(df['Sex'],np.zeros_like(df['Sex']))
plt.plot(df['Sex'],np.zeros_like(df['Sex']))
plt.plot(df['Sex'],np.zeros_like(df['Sex']))
plt.show()
```



Bi-Variate Analysis

```
In [13]: sns.factorplot(x=hue="Rings",size=5, map=plt.scatter,"Length","Weight",add_legend=)
```

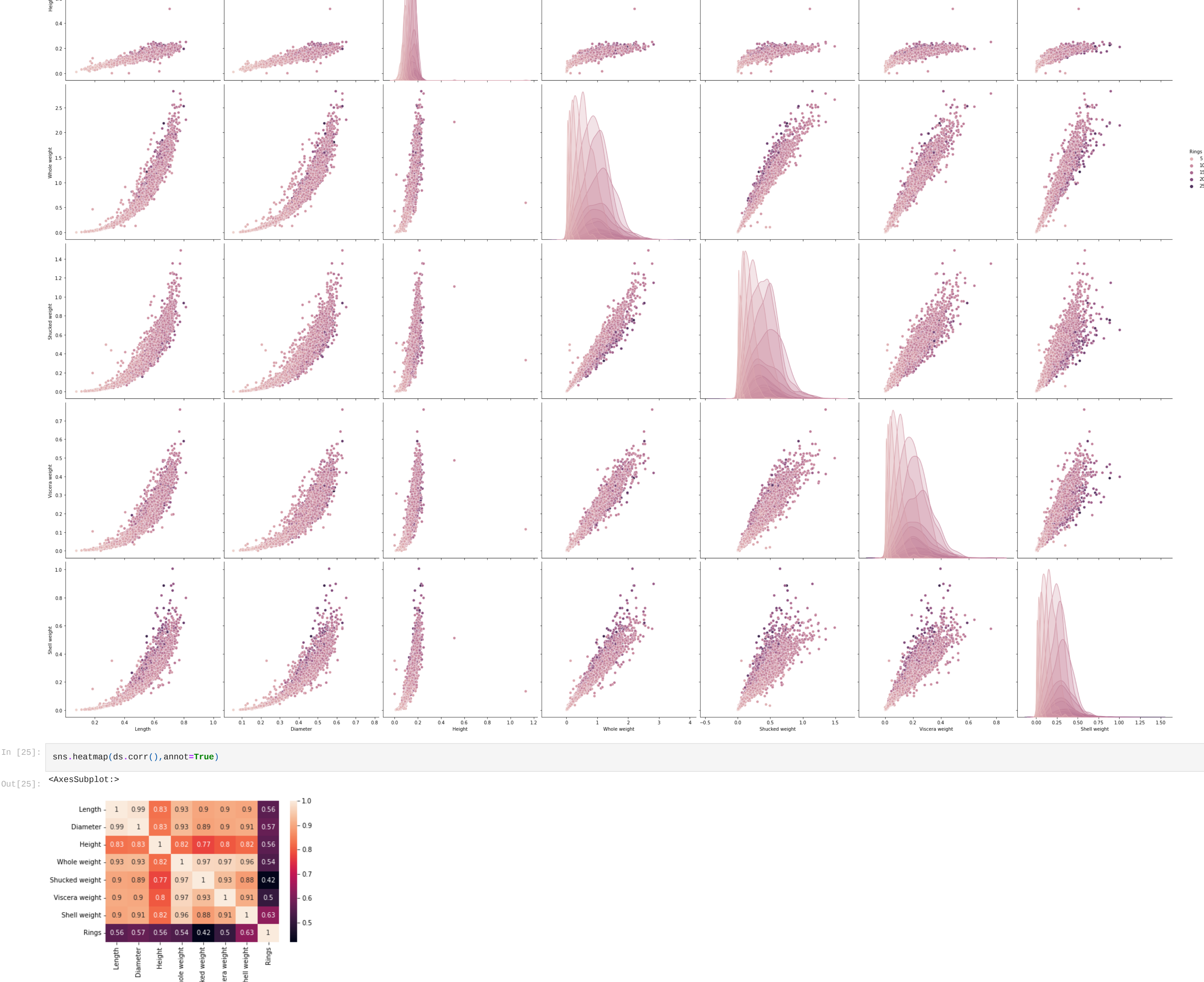
C:\Users\jagadeesan\Anaconda3\lib\site-packages\seaborn\axisgrid.py:337: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



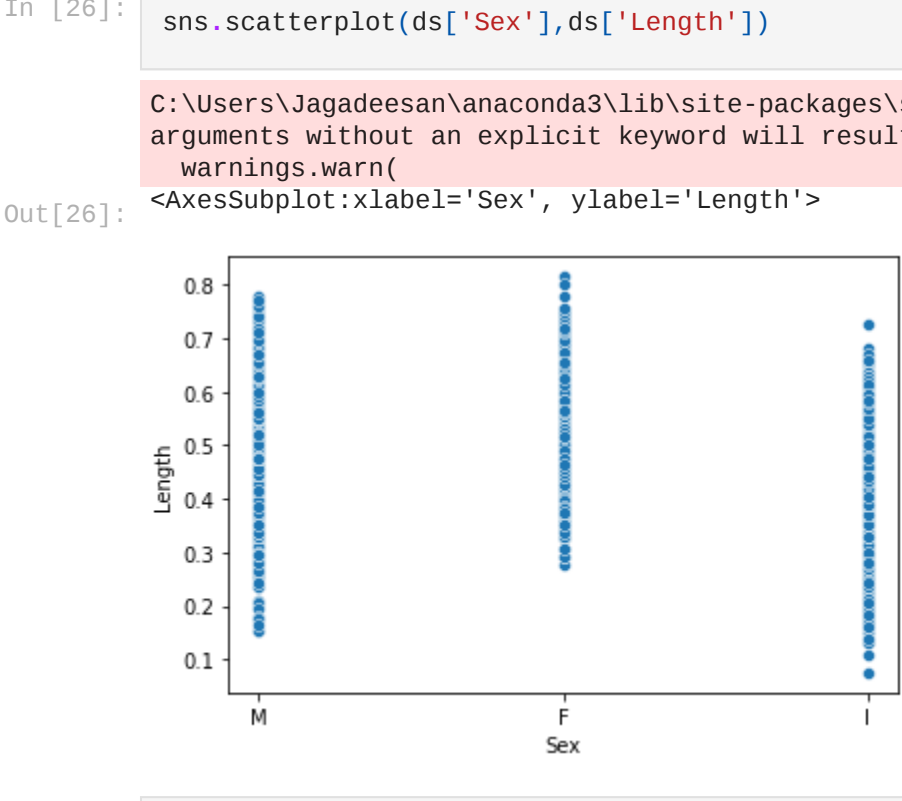
Multi-Variate Analysis

```
In [14]: sns.pairplot(ds,hue="Rings",size=5)
```

C:\Users\jagadeesan\Anaconda3\lib\site-packages\seaborn\axisgrid.py:2676: UserWarning: The 'size' parameter has been renamed to 'height'; please update your code.
warnings.warn(msg, UserWarning)



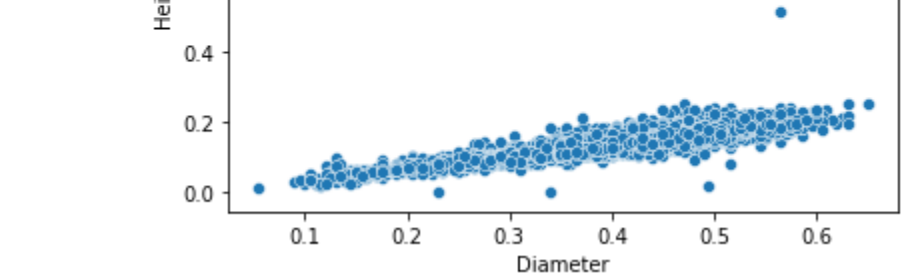
```
In [25]: sns.heatmap(ds.corr(),annot=True)
```



```
In [26]: sns.scatterplot(ds['Sex'],ds['Length'])
```

C:\Users\jagadeesan\Anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(msg, FutureWarning)

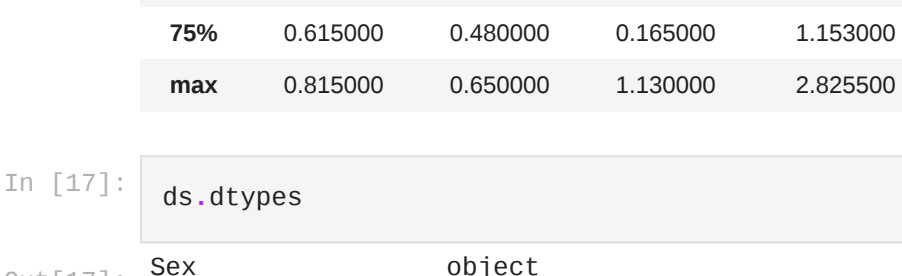
```
Out[26]: <AxesSubplot:label='Sex', ylabel='Length'>
```



```
In [27]: sns.scatterplot(ds['Diameter'],ds['Height'])
```

C:\Users\jagadeesan\Anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(msg, FutureWarning)

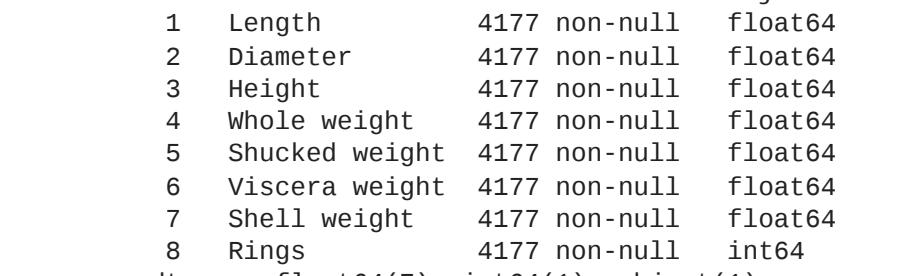
```
Out[27]: <AxesSubplot:label='Diameter', ylabel='Height'>
```



```
In [27]: sns.scatterplot(ds['Diameter'],ds['Height'])
```

C:\Users\jagadeesan\Anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be 'data', and passing other arguments without an explicit keyword will result in an error or misinterpretation.
warnings.warn(msg, FutureWarning)

```
Out[27]: <AxesSubplot:label='Diameter', ylabel='Height'>
```



```
In [1]: # descriptive statistics
```

```
In [16]: ds.describe()
```

```
Out[16]:
```

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
count	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000
mean	0.523992	0.407881	0.139516	0.928742	0.395967	0.189584	0.238831	9.932684
std	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0.126203	3.224169
min	0.075000	0.055000	0.000000	0.002000	0.001000	0.005000	0.001000	0.000000
25%	0.450000	0.350000	0.115000	0.441000	0.160000	0.095000	0.130000	6.000000
50%	0.500000	0.400000	0.140000	0.766000	0.300000	0.170000	0.230000	9.000000
75%	0.615000	0.480000	0.160000	1.153000	0.500000	0.263000	0.290000	11.000000
max	0.815000	0.650000	0.130000	2.855000	1.488000	0.760000	1.095000	29.000000

```
In [17]: ds.dtypes
```

```
Out[17]:
```

	Sex	object
Length	float64	
Diameter	float64	
Height	float64	
Whole weight	float64	
Shucked weight	float64	
Viscera weight	float64	
Shell weight	float64	
Rings	int64	
dtype: object		

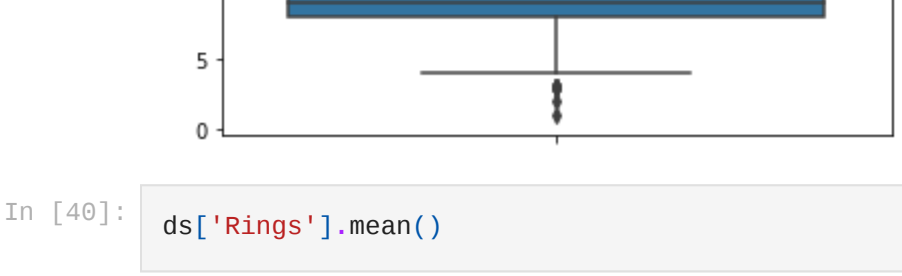
```
In [29]: ds.info()
```

```
<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
dtype: object
memory usage: 293.8 KB
```

```
In [1]: #outliers
```

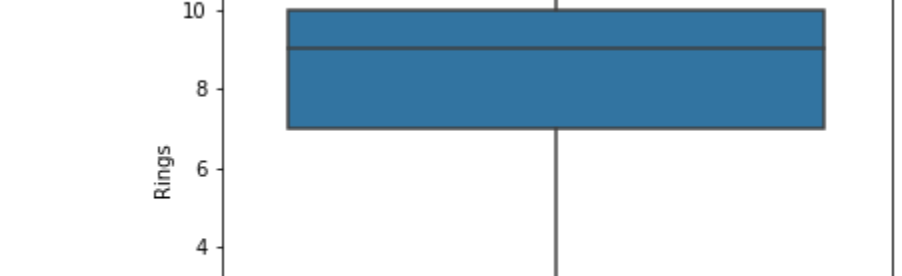
```
In [34]: sns.displot(ds['Sex'])
```

```
Out[34]: <seaborn.axisgrid.FacetGrid at 0x2989885780>
```



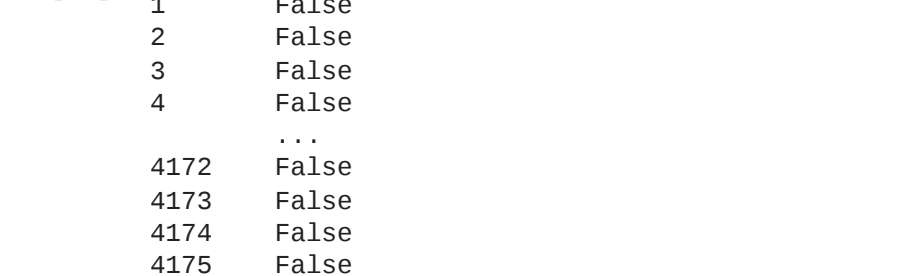
```
In [35]: sns.boxplot(x='Sex',y='Rings',data=ds)
```

```
Out[35]: <AxesSubplot:label='Sex', ylabel='Rings'>
```



```
In [36]: sns.boxplot(y='Rings',data=ds)
```

```
Out[36]: <AxesSubplot:ylabel='Rings'>
```



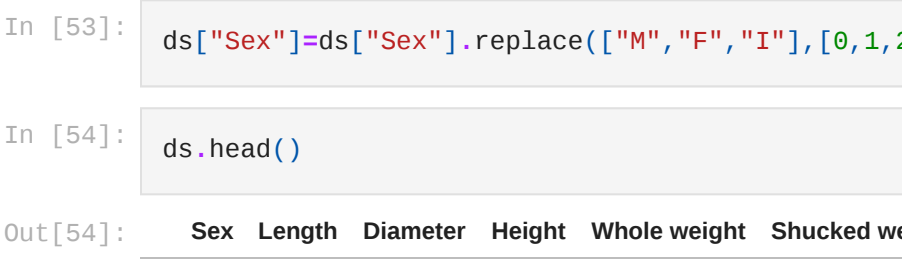
```
Out[36]: ds['Rings'].mean()
```

```
Out[36]: 9.932684462532915
```

```
Out[40]: data=ds[ds['Rings']>12]
```

```
In [45]: sns.boxplot(y='Rings',data=data)
```

```
Out[45]: <AxesSubplot:ylabel='Rings'>
```



missing values

```
In [48]: ds.duplicated()
```

```
Out[48]:
```

	0	False
1	False	
2	False	
3	False	
4	False	
...	...	
4172	False	
4173	False	
4174	False	
4175	False	
4176	False	

```
Length: 4177, dtype: bool
```

```
In [49]: ds.isna().any().sum()
```

```
Out[49]: 0
```

```
In [50]: ds['Sex'].unique()
```

```
Out[50]: array(['M', 'F', 'I'], dtype=object)
```

```
In [51]: from sklearn.preprocessing import LabelEncoder
```

```
In [52]: le=LabelEncoder
```

```
In [53]: ds['Sex']=ds['Sex'].replace({'M':'0','F':'1','I':'2'})
```

```
In [54]: ds.head()
```

```
Out[54]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	0	0.455	0.365	0.095	0.5140	0.2245	0.1050	0.100	15
1	0	0.360	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	1	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	0	0.440	0.365	0.125	0.5160	0.2195	0.1140	0.155	10
4	2	0.320	0.255	0.080	0.2260	0.0995	0.0295	0.095	7

```
In [55]: ds.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
dtype: object
memory usage: 293.8 KB
```

```
In [56]: ds.describe()
```

```
Out[56]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
count	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000	4177.00000
mean	0.955470	0.523992	0.407881	0.139516	0.928742	0.395967	0.189584	0.238831	9.932684
std	0.827155	0.120093	0.099240	0.041827	0.490389	0.221963	0.109614	0.126203	3.224169
min	0.000000	0.075000	0.055000	0.000000	0.002000	0.001000	0.005000	0.001000	0.000000
25%	1.000000	0.450000	0.350000	0.115000	0.441000	0.160000	0.095000	0.130000	6.000000
50%	1.000000	0.540000	0.425000	0.140000	0.766500	0.338000	0.171000	0.234000	9.000000
75%	2.000000	0.615000	0.480000	0.160000	1.153000	0.502000	0.263000	0.290000	11.000000
max	2.000000	0.815000	0.650000	0.130000	2.855000	1.488000	0.760000	1.095000	29.000000

```
In [57]: ds.duplicated().sum()
```

```
Out[57]: 0
```

```
In [58]: ds.corr()
```

```
Out[58]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
Sex	1.000000	-0.448705	-0.454565	-0.417628	-0.461138	-0.440927	-0.454658	-0.445549	-0.318027
Length	-0.448705	1.000000	0.986812	0.827584	0.923261	0.897924	0.930533	0.897706	0.556720
Diameter	-0.454565	0.986812	1.000000	0.839884	0.925452	0.899122	0.899724	0.903300	0.574660
Height	-0.417628	0.827584	0.839884	1.000000	0.816725	0.716972	0.780359	0.817328	0.556767
Whole weight	-0.461138	0.923261	0.925452	0.816725	1.000000	0.959495	0.963375	0.953556	0.540894
Shucked weight	-0.440927	0.897924	0.899122	0.716972	0.959495	1.000000	0.931161	0.882617	0.426984
Viscera weight	-0.454658	0.930533	0.899724	0.780359	0.963375	0.931161	1.000000	0.967696	0.502819
Shell weight	-0.445549	0.897706	0.903300	0.817328	0.953556	0.882617	0.967696	1.000000	0.627574
Rings	-0.318027	0.556720	0.574660	0.556767	0.540894	0.426984	0.502819	0.627574	1.000000

split of dependent and independent value

```
In [59]: x=ds.drop('Rings',axis=1)
```

```
Out[59]: y=ds['Rings']
```

```
In [62]: x
```

```
Out[62]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
0	0	0.455	0.365	0.095	0.5140	0.2245	0.1050	0.100
1	0	0.360	0.265	0.090	0.2255	0.0995	0.0485	0.070
2	1	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210
3	0	0.440	0.365	0.125	0.5160	0.2195	0.1140	0.155
4	2	0.320	0.255	0.080	0.2260	0.0995	0.0295	0.095

```
Out[62]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
4172	1	0.565	0.460	0.165	0.8870	0.3760	0.2360	0.240
4173	0	0.560	0.440	0.155	0.9060	0.4300	0.2145	0.265
4174	0	0.800	0.475	0.205	1.1760	0.5355	0.3875	0.300
4175	1	0.625	0.485	0.150	1.0845	0.5310	0.2610	0.290
4176	0	0.710	0.555	0.195	1.9465	0.9455	0.3765	0.450

```
4177 rows x 8 columns
```

```
Out[63]: y
```

```
Out[63]:
```

	0	15	7	9	10	11
0	15					
1	7					
2	9					
3	10					
4	7					
...
4172	11					
4173	10					
4174	9					
4175	10					
4176	12					

```
In [64]: data=ds.drop('dummies=ds').data
```

```
Out[64]:
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	0	0.455	0.365	0.095	0.5140	0.2245	0.1050	0.100	15
1	0	0.360	0.265	0.090	0.2255	0.0995	0.0485	0.070	7
2	1	0.530	0.420	0.135	0.6770	0.2565	0.1415	0.210	9
3	0	0.440	0.365	0.125	0.5160	0.2195	0.1140	0.155	10
4	2	0.320	0.255	0.080	0.2260	0.0995	0.0295	0.095	7

```
Out[64]:
```

	Sex	Length
--	-----	--------