Assignment -4

Python Programming

Assignment Date	4 November 2022
Student Name	Mr. N. Tamizh selvan
Student Roll Number	621319205303
Maximum Marks	2 Marks

Questions:

```
import numpy as np
import seaborn as sns
import pandas as pd
from matplotlib import pyplot as plt
```

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

```
Mounted at /content/drive
```

```
data=pd.read_csv('/content/drive/MyDrive/Details/IBM/abalone.csv')
```

```
data
```

	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

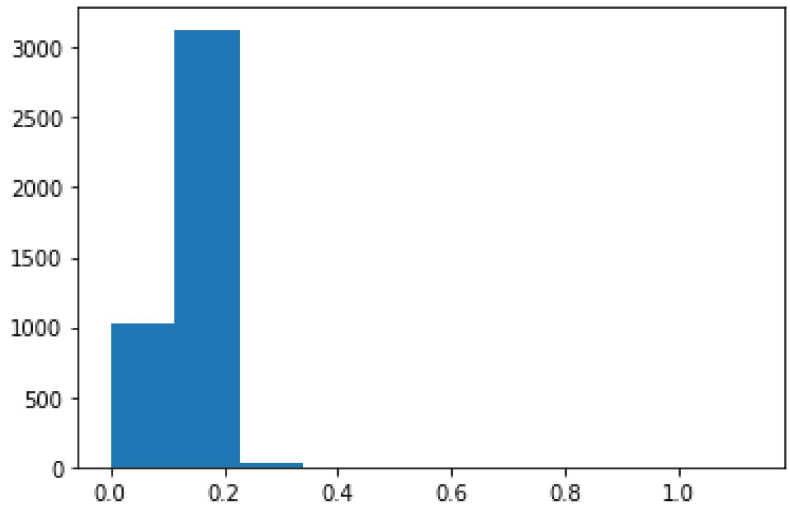
```
data.head(5)
```

```
data.tail(5)
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rir
4172	F	0.565	0.450	0.165	0.8870	0.3700	0.2390	0.2490	
4173	M	0.590	0.440	0.135	0.9660	0.4390	0.2145	0.2605	
4174	M	0.600	0.475	0.205	1.1760	0.5255	0.2875	0.3080	
4175	F	0.625	0.485	0.150	1.0945	0.5310	0.2610	0.2960	

```
plt.hist(data['Height'])
```

```
(array([1.023e+03, 3.129e+03, 2.300e+01, 0.000e+00, 1.000e+00, 0.000e+00,
        0.000e+00, 0.000e+00, 0.000e+00, 1.000e+00]),
 array([0.   , 0.113, 0.226, 0.339, 0.452, 0.565, 0.678, 0.791, 0.904,
        1.017, 1.13 ]),
 <a list of 10 Patch objects>)
```




```
plt.boxplot(data['Height'])
```

```
{'whiskers': [<matplotlib.lines.Line2D at 0x7f825da15e50>,
  <matplotlib.lines.Line2D at 0x7f825da1d3d0>],
 'caps': [<matplotlib.lines.Line2D at 0x7f825da1d910>,
  <matplotlib.lines.Line2D at 0x7f825da1de50>],
 'boxes': [<matplotlib.lines.Line2D at 0x7f825da15990>],
 'medians': [<matplotlib.lines.Line2D at 0x7f825da22410>],
 'fliers': [<matplotlib.lines.Line2D at 0x7f825da22950>],
 'means': []}
```

```
data.isnull()
```

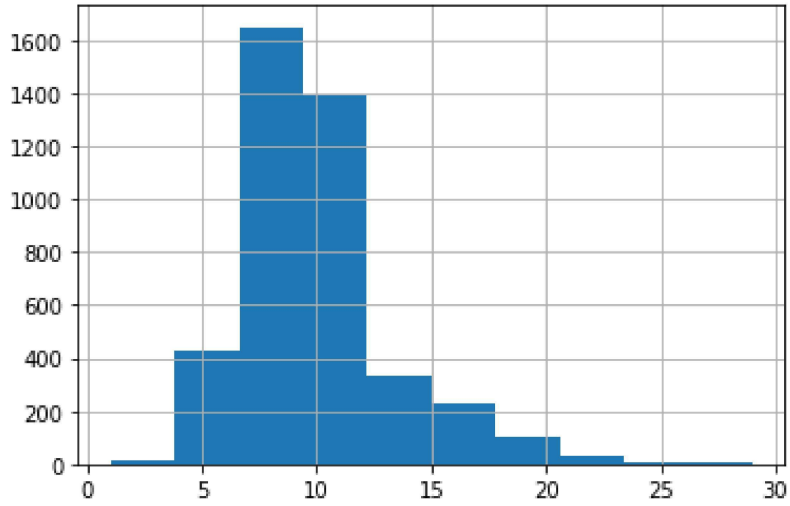
	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	R
0	False	False	False	False	False	False	False	False	F
1	False	False	False	False	False	False	False	False	F
2	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	F
4	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	
4172	False	False	False	False	False	False	False	False	F
4173	False	False	False	False	False	False	False	False	F
4174	False	False	False	False	False	False	False	False	F
4175	False	False	False	False	False	False	False	False	F

```
data.notnull()
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Rings
0	True	True	True	True	True	True	True	True	True

```
data['Rings'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f825d9a1bd0>
```

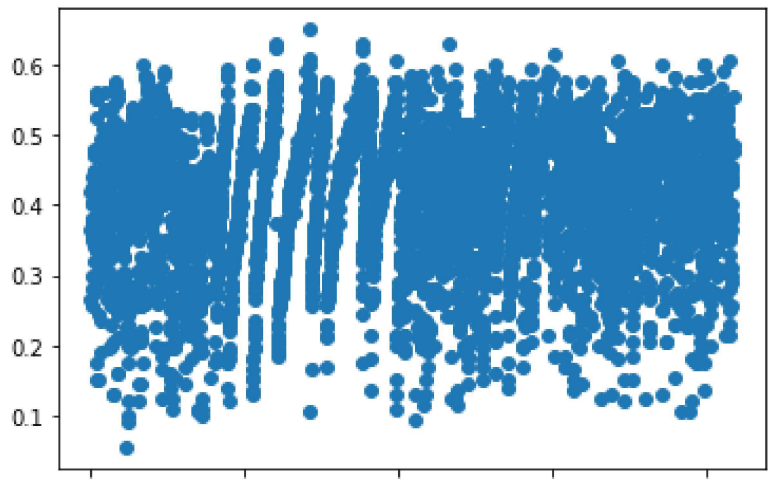



```
data.dtypes
```

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

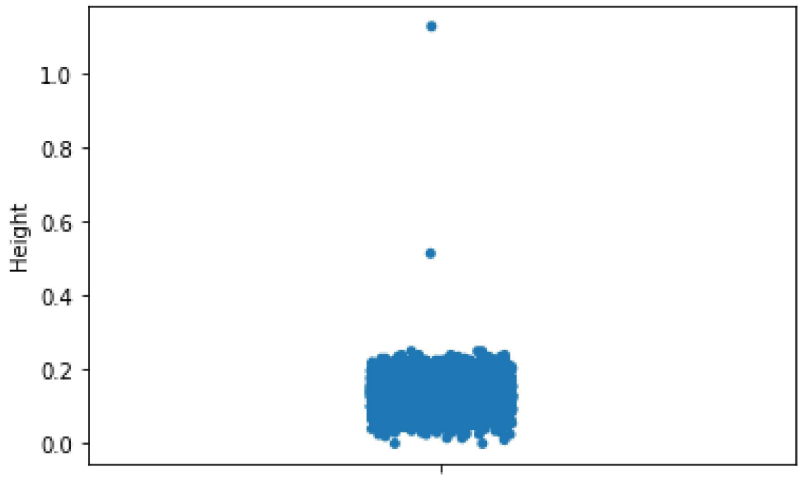
```
plt.scatter(data.index,data['Diameter'])
```

```
<matplotlib.collections.PathCollection at 0x7f825d87fcd0>
```



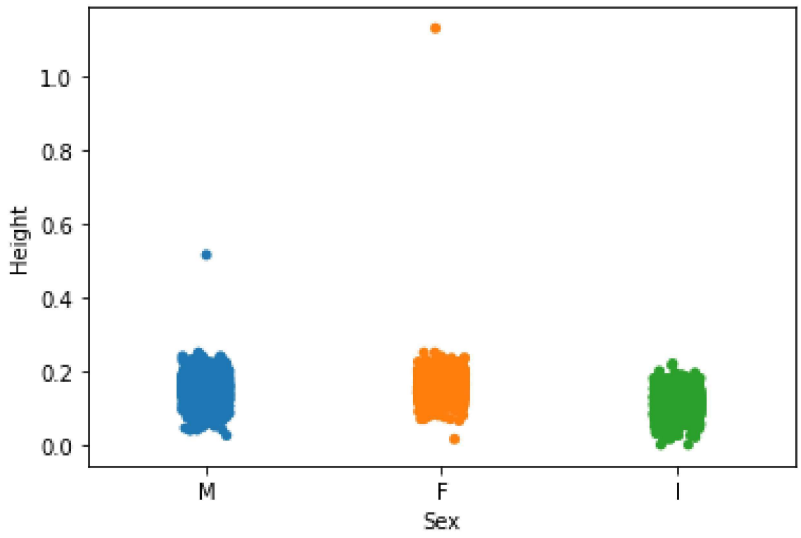
```
sns.stripplot(y=data['Height'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f825d831310>
```




```
sns.stripplot(x=data['Sex'],y=data['Height'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f825d7aa790>
```

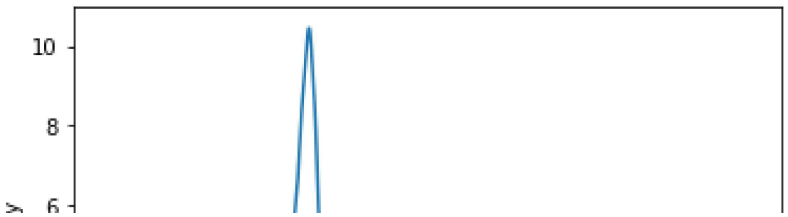



```
plt.figure(figsize=(5,5))
```

```
<Figure size 360x360 with 0 Axes>
<Figure size 360x360 with 0 Axes>
```

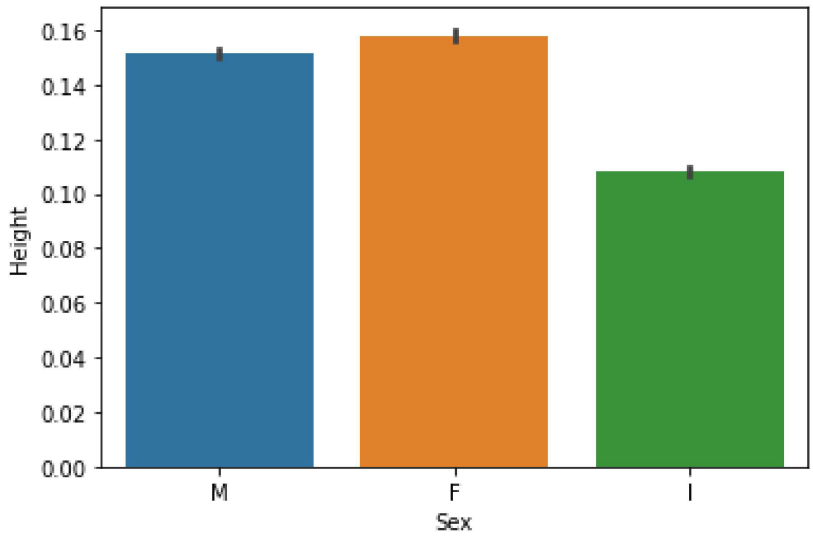
```
data['Height'].plot(kind='density')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f825d7a3350>
```



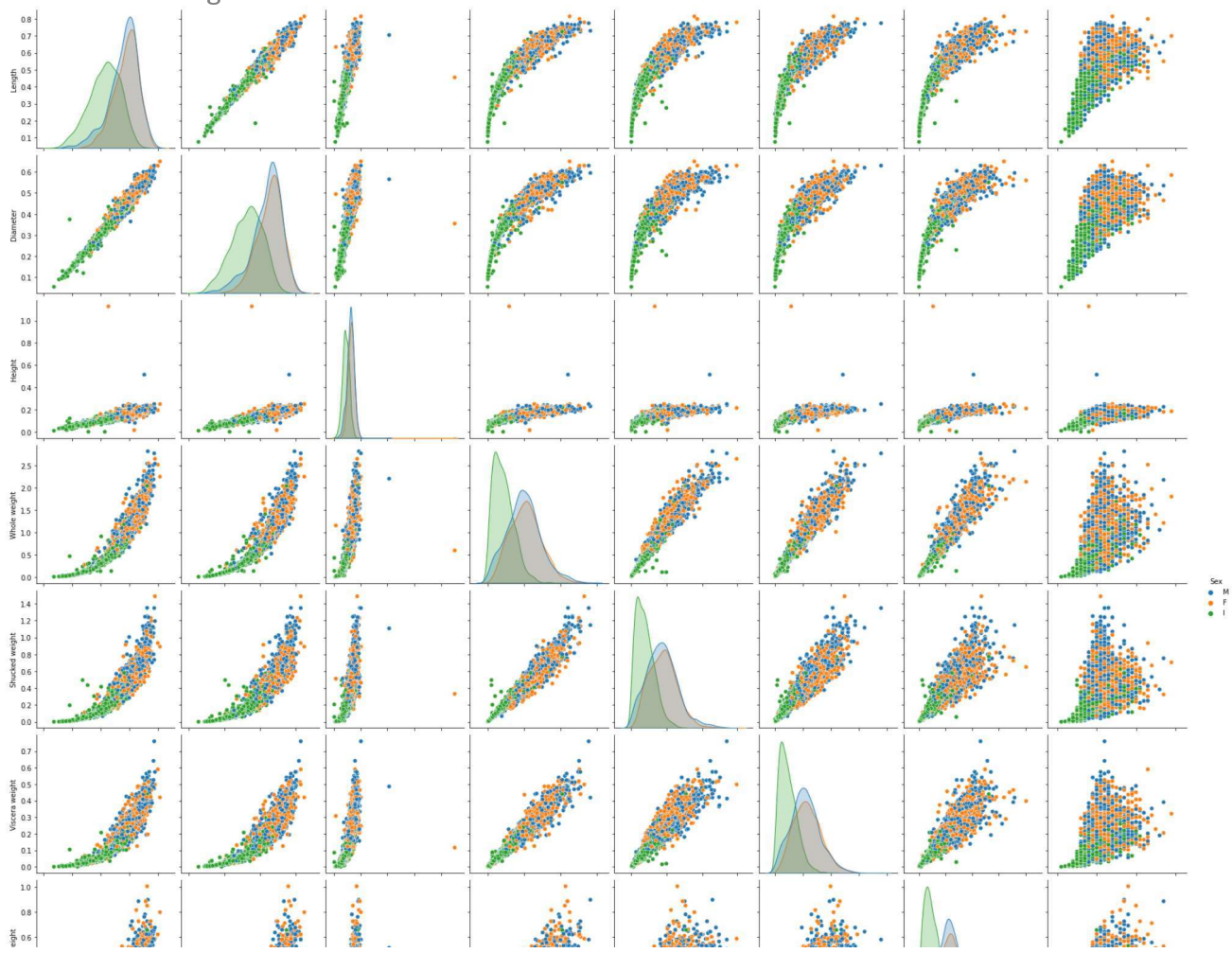

```
sns.barplot(x='Sex',y='Height',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f825d71dcd0>
```



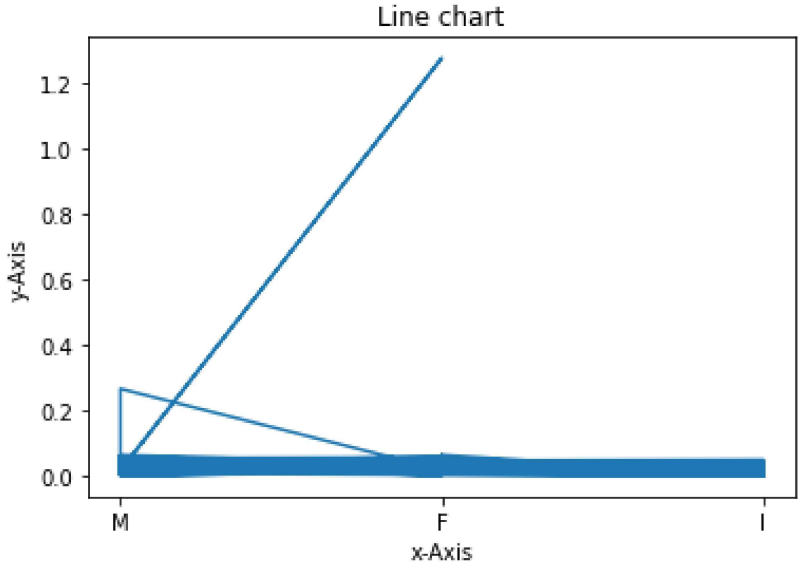

```
sns.pairplot(data,hue="Sex",size=3)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: Th
  warnings.warn(msg, UserWarning)
<seaborn.axisgrid.PairGrid at 0x7f825d625590>
```



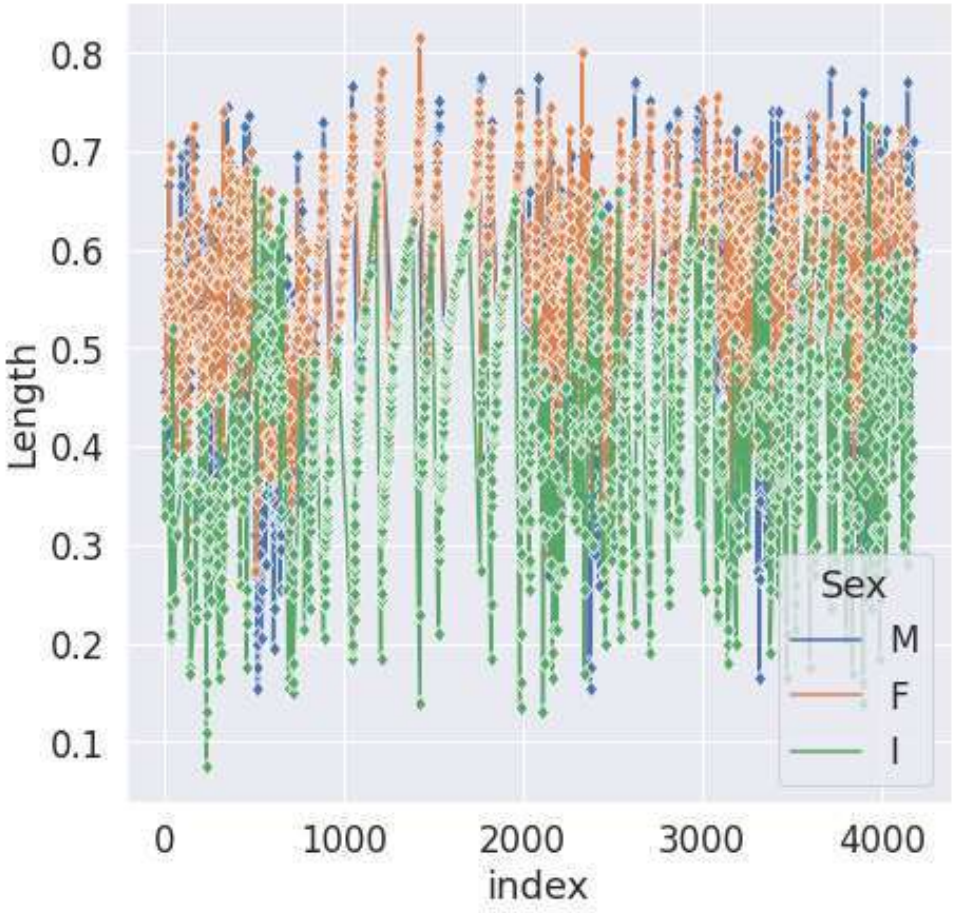

```
x = np.array(data['Sex'])
y = np.power(data['Height'],2)
plt.plot(x,y)
plt.title("Line chart")
plt.xlabel("x-Axis")
plt.ylabel("y-Axis")
```

```
Text(0, 0.5, 'y-Axis')
```



```
sns.set(rc={'figure.figsize': (7,7)})
sns.set (font_scale=1.5)
fig=sns.lineplot (x=data.index, y=data['Length'], markevery=1, marker='d', data=data,
fig.set(xlabel='index')
```

```
[Text(0.5, 0, 'index')]
```



TRAIN AND TEST:

```
from sklearn.model_selection import train_test_split
```

```
X=data.iloc[ : , :-1]
y=data.iloc[:, -1]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state
X_train
```

	Sex	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight
678	F	0.450	0.380	0.165	0.8165	0.2500	0.1915	0.2650
3009	I	0.255	0.185	0.065	0.0740	0.0305	0.0165	0.0200
1906	I	0.575	0.450	0.135	0.8245	0.3375	0.2115	0.2390
768	F	0.550	0.430	0.155	0.7850	0.2890	0.2270	0.2330
2781	M	0.595	0.475	0.140	1.0305	0.4925	0.2170	0.2780

```
y_train
```

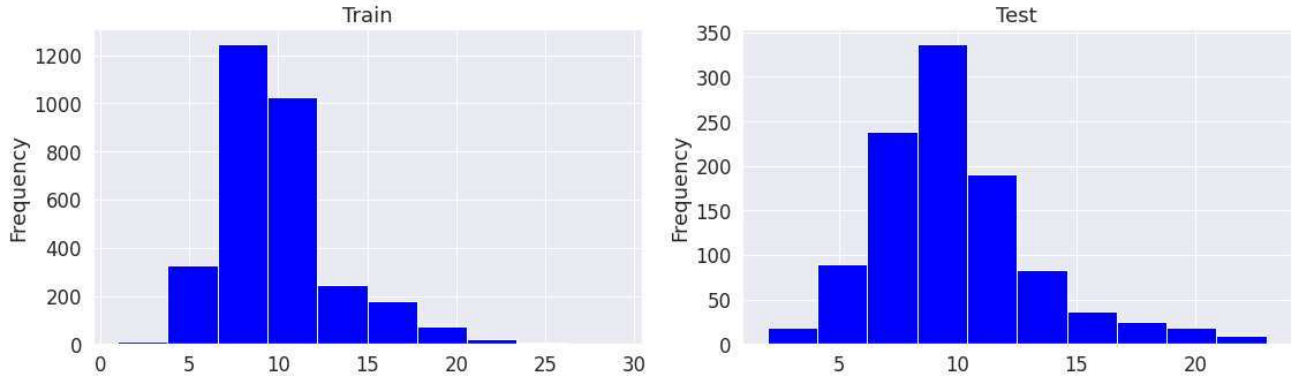
```
678     23
3009     4
1906    11
768     11
2781    10
        ..
1033    10
3264    12
1653    10
2607     9
2732     8
Name: Rings, Length: 3968, dtype: int64
```

```
train, test = train_test_split(data, test_size=0.25, random_state=1)
print('Train data points :', len(train))
print('Test data points :', len(test))
```

```
Train data points : 3132
Test data points : 1045
```

```
train.Sex = train.Sex.replace({"M":1, "I":0, "F":-1})
test.Sex = test.Sex.replace({"M":1, "I":0, "F":-1})
numerical_features = ["Length", 'Diameter', 'Height','Whole weight','Shucked weight',
categorical_feature = "Sex"
features = numerical_features + [categorical_feature]
target = 'Rings'
```

```
fig, axes = plt.subplots(ncols=2,figsize=(16, 5))
train[target].plot.hist(color='blue', ax=axes[0])
axes[0].set(title="Train")
test[target].plot.hist(color='blue', ax=axes[1])
axes[1].set(title="Test")
plt.tight_layout()
plt.show()
```




```
X_train = train[features]
y_train = train[target]
X_test = test[features]
y_test = test[target]
X_train.head()
```

	Length	Diameter	Height	Whole weight	Shucked weight	Viscera weight	Shell weight	Sex
4014	0.625	0.480	0.175	1.0650	0.4865	0.2590	0.285	1
3252	0.480	0.380	0.130	0.6175	0.3000	0.1420	0.175	1
305	0.200	0.145	0.060	0.0370	0.0125	0.0095	0.011	0
1857	0.505	0.400	0.145	0.7045	0.3340	0.1425	0.207	0
439	0.500	0.415	0.165	0.6885	0.2490	0.1380	0.250	1

```
from sklearn.compose import make_column_selector as selector
categorical_columns_selector = selector(dtype_include=object)
categorical_columns = categorical_columns_selector(data)
categorical_columns
```

```
['Sex']
```

```
data_categorical = data[categorical_columns]
data_categorical.head()
```

	Sex
0	M
2	I
3	M
4	I

Double-click (or enter) to edit