In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import statsmodels.api as sm
from scipy.stats import pearsonr
import seaborn as sns
```

In [2]:
```
data=pd.read_csv("C:/Users/ADMIN/Documents/weight-height.csv",na_values="?")
```

In [3]:
```
data.head()
```

Out[3]:

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801

In [4]:
```
data.value_counts()
```

Out[4]:
```
Gender  Height     Weight
Female  54.263133  64.700127     1
Male    67.030001  172.776569    1
        67.889015  179.934465    1
        67.910251  170.457826    1
        67.914286  192.066392    1
                                ..
Female  64.854997  146.692798    1
        64.855262  150.397118    1
        64.859973  140.784838    1
        64.860659  159.568226    1
Male    78.998742  269.989698    1
Length: 10000, dtype: int64
```

In [5]:
```
data.shape
```

Out[5]:
```
(10000, 3)
```

In [6]:
```
data.columns
```

Out[6]:
```
Index(['Gender', 'Height', 'Weight'], dtype='object')
```

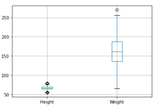
In [7]:
```
data.info
```

Out[7]:
```
<bound method DataFrame.info of       Gender     Height      Weight
0       Male  73.847017  241.893563
1       Male  68.781904  162.310473
2       Male  74.110105  212.740856
3       Male  71.730978  220.042470
4       Male  69.881796  206.349801
...      ...        ...         ...
9995  Female  66.172652  136.777454
9996  Female  67.067155  170.867906
9997  Female  63.867992  128.475319
9998  Female  69.034243  163.852461
9999  Female  61.944246  113.649103

[10000 rows x 3 columns]>
```

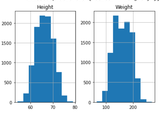
In [8]:
```
data.boxplot()
```

Out[8]:
```
<AxesSubplot:>
```



In [9]:
```
data.hist()
```

Out[9]:
```
array([[<AxesSubplot:title={'center':'Height'}>,
        <AxesSubplot:title={'center':'Weight'}>]], dtype=object)
```



In [17]:
```
data.isnull().sum()
```

Out[17]:
```
Gender    0
Height    0
Weight    0
dtype: int64
```

In [11]:
```
data.Gender.describe()
```

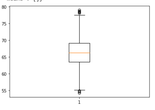
Out[11]:
```
count     10000
unique        2
top        Male
freq       5000
Name: Gender, dtype: object
```

In [12]:
```
data.Gender.value_counts()
```

Out[12]:
```
Male      5000
Female    5000
Name: Gender, dtype: int64
```

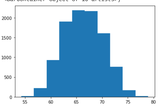
In [13]:
```
plt.boxplot(data.Height)
```

Out[13]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x15df52af250>,
  <matplotlib.lines.Line2D at 0x15df52af5e0>],
 'caps': [<matplotlib.lines.Line2D at 0x15df52af970>,
  <matplotlib.lines.Line2D at 0x15df52afd00>],
 'boxes': [<matplotlib.lines.Line2D at 0x15df529fe80>],
 'medians': [<matplotlib.lines.Line2D at 0x15df52ba890>],
 'fliers': [<matplotlib.lines.Line2D at 0x15df52ba460>],
 'means': []}
```



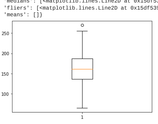
In [14]:
```
plt.hist(data.Height)
```

Out[14]:
```
(array([  23.,  218.,  926., 1906., 2196., 2167., 1612.,  765.,  163.,
         24.]),
 array([54.26313333, 56.73569423, 59.21025513, 61.68381604, 64.15737694,
        66.63093784, 69.10449874, 71.57805964, 74.05162055, 76.52518145,
        78.99874235]),
 <BarContainer object of 10 artists>)
```



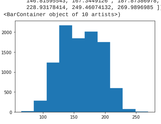
In [15]:
```
plt.boxplot(data.Weight)
```

Out[15]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x15df537df70>,
  <matplotlib.lines.Line2D at 0x15df538d340>],
 'caps': [<matplotlib.lines.Line2D at 0x15df538d6d0>,
  <matplotlib.lines.Line2D at 0x15df538da60>],
 'boxes': [<matplotlib.lines.Line2D at 0x15df537dbe0>],
 'medians': [<matplotlib.lines.Line2D at 0x15df538df70>],
 'fliers': [<matplotlib.lines.Line2D at 0x15df539a1c0>],
 'means': []}
```



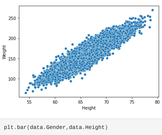
In [16]:
```
plt.hist(data.Weight)
```

Out[16]:
```
(array([  15.,  278.,  1241.,  2175.,  1849.,  2016.,  1747.,   598.,    71.,
          10.]),
 array([ 64.70012671,  85.22908389, 105.75804107, 126.28699825,
        146.81595543, 167.3449126 , 187.87386978, 208.40282696,
        228.93178414, 249.46074132, 269.9896985 ]),
 <BarContainer object of 10 artists>)
```



In [18]:
```
sns.scatterplot(x="Height",y="Weight",data=data)
```

Out[18]:
```
<AxesSubplot:xlabel='Height', ylabel='Weight'>
```



In [22]:
```
plt.bar(data.Gender,data.Height)
```

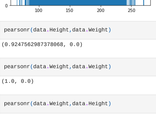
Out[22]:
```
<BarContainer object of 10000 artists>
```

In [23]:
```
plt.bar(data.Gender,data.Weight)
```

Out[23]:
```
<BarContainer object of 10000 artists>
```

In [24]:
```
plt.bar(data.Weight,data.Height)
```

Out[24]:
```
<BarContainer object of 10000 artists>
```



In [26]:
```
pearsonr(data.Height,data.Weight)
```

Out[26]:
```
(0.9247562987378068, 0.0)
```

In [28]:
```
pearsonr(data.Weight,data.Weight)
```

Out[28]:
```
(1.0, 0.0)
```

In [30]:
```
pearsonr(data.Weight,data.Height)
```

Out[30]:
```
(0.9247562987378068, 0.0)
```

In [ ]:
```

```

In [ ]:
```

```