```
In [3]:  import pandas as pd
         import numpy as np
         import matplotlib.pyplot as plt
```

```
In [ ]:
```

```
In [7]:  my_data1=pd.read_excel(r"DataSets/Lab1data.xlsx")
         my_data1.shape
```

Out[7]: (12, 9)

```
In [ ]:
```

```
In [8]:  my_data1.columns
```

Out[8]: Index(['Stu ID', 'Gender', 'Age', 'Quantitative', 'Verbal', 'Technical',
               'Interview', 'Total', 'Results'],
              dtype='object')

```
In [9]:  my_data1.head(2)
```

Out[9]:

	Stu ID	Gender	Age	Quantitative	Verbal	Technical	Interview	Total	Results
0	A892	Male	31.0	22	33.0	35	29	119	0
1	A985	Male	28.0	46	55.0	70	70	241	1

```
In [ ]:
```

```
In [ ]:
```

```
In [10]:  my_data1.dtypes
```

```
Out[10]:  Stu ID           object
          Gender           object
          Age             float64
          Quantitative      int64
          Verbal          float64
          Technical         int64
          Interview         int64
          Total             int64
          Results           int64
          dtype: object
```

In [ ]:

In [11]: `my_data1.describe()`

Out[11]:

|	Age	Quantitative	Verbal	Technical	Interview	Total	Results
count	11.000000	12.000000	11.000000	12.000000	12.000000	12.000000	12.000000
mean	30.090909	45.000000	50.727273	48.500000	41.166667	186.916667	0.416667
std	2.300198	17.331002	15.060484	13.304135	17.708926	37.497778	0.514929
min	27.000000	22.000000	23.000000	30.000000	22.000000	119.000000	0.000000
25%	28.000000	34.250000	42.500000	37.000000	25.500000	172.000000	0.000000
50%	30.000000	44.500000	55.000000	47.000000	38.500000	188.000000	0.000000
75%	31.500000	56.000000	62.000000	57.500000	52.250000	209.500000	1.000000
max	34.000000	72.000000	69.000000	70.000000	70.000000	241.000000	1.000000

In [ ]:

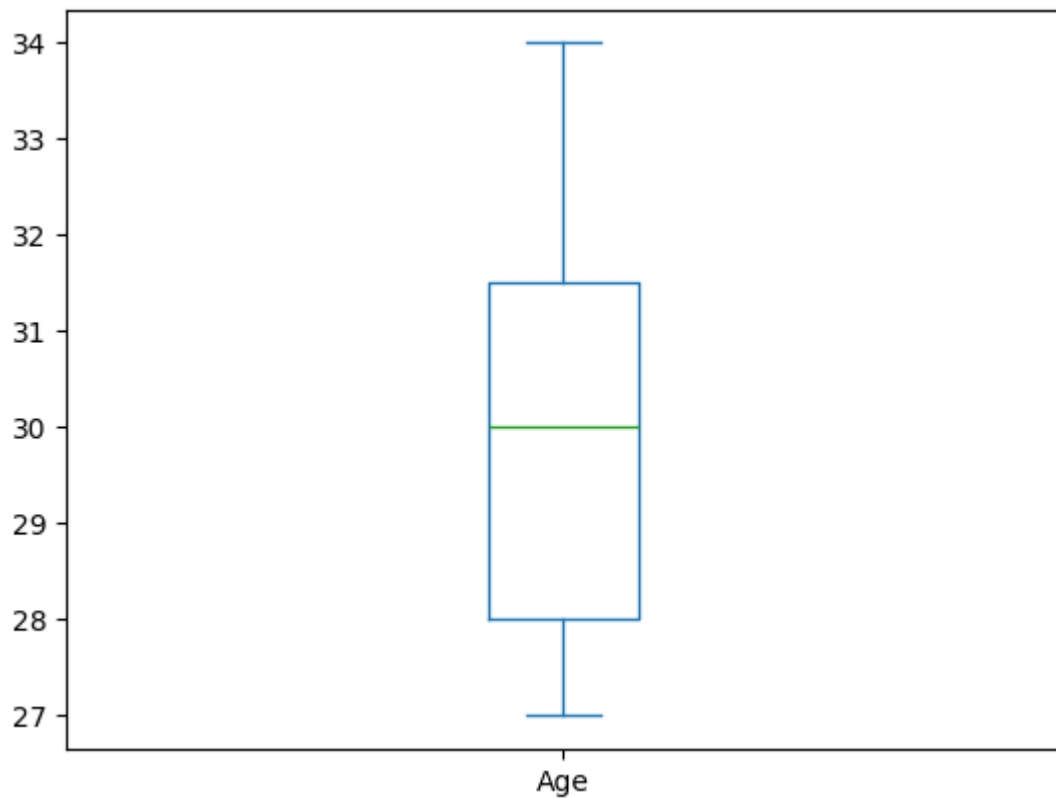In [12]: `my_data1['Age'].plot.hist(bins=6)`

Out[12]: `<Axes: ylabel='Frequency'>`

In [ ]:
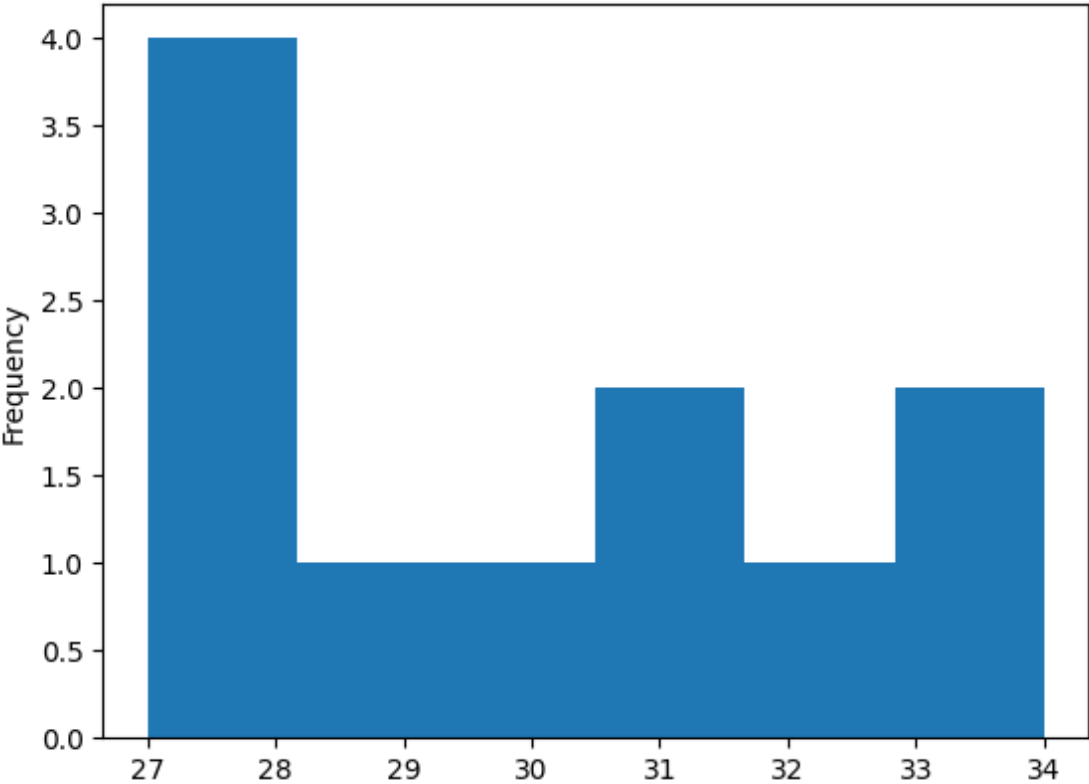
In [13]:
```python
my_data1['Age'].plot.box()
```

Out[13]: <Axes: >

In [ ]:

In [14]:
```python
my_data1['Gender'].value_counts()
```
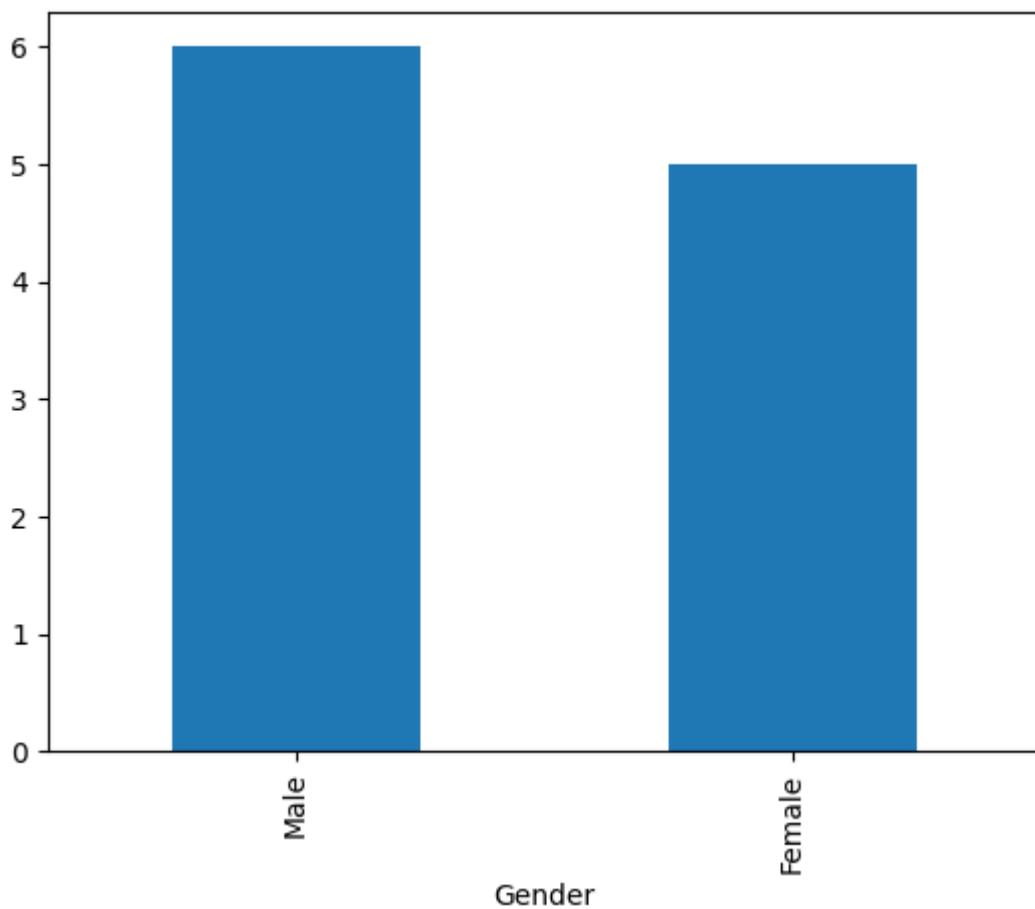
Out[14]:
```
Gender
Male      6
Female    5
Name: count, dtype: int64
```

In [ ]:

In [15]:
```python
my_data1['Gender'].value_counts().plot.bar()
```

Out[15]: `<Axes: xlabel='Gender'>`

In [ ]:

In [16]:
```python
my_data1['Gender'].value_counts()/len(my_data1['Gender'])
```
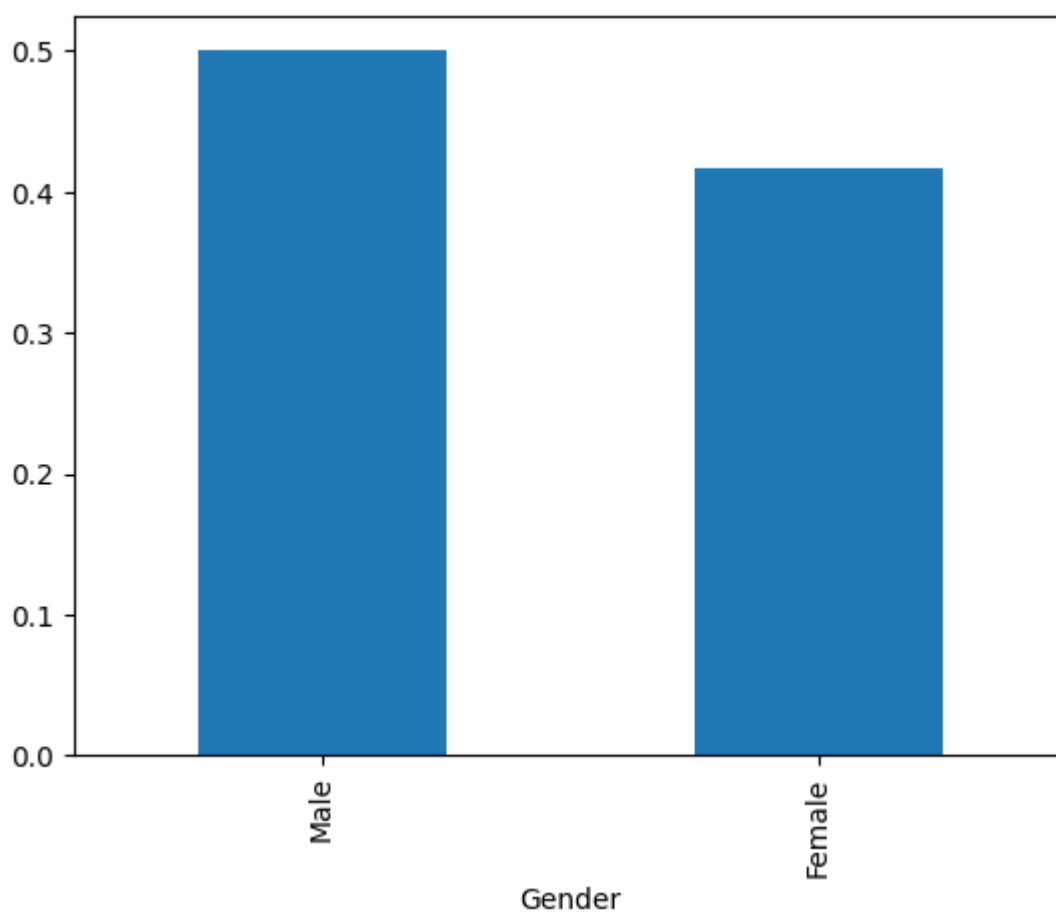
Out[16]:
```
Gender
Male       0.500000
Female     0.416667
Name: count, dtype: float64
```

In [ ]:

```
In [17]: (my_data1['Gender'].value_counts()/len(my_data1['Gender'])).plot.bar()
```
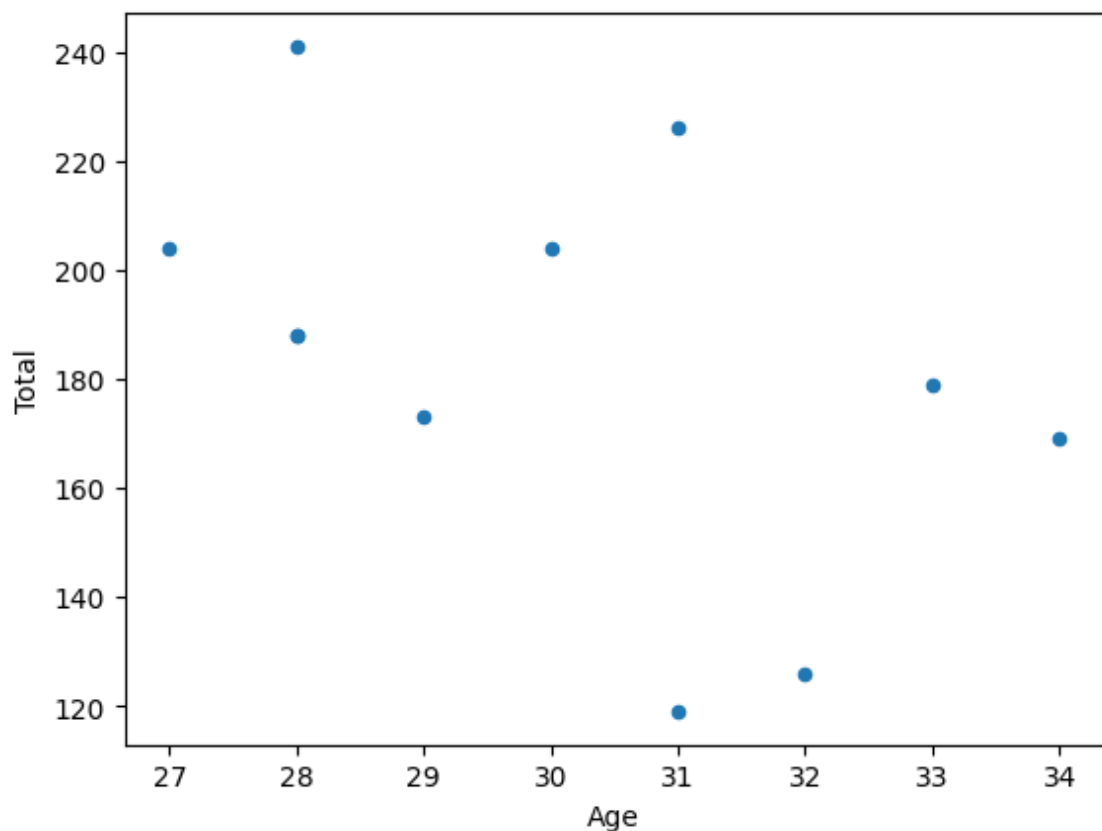
Out[17]: <Axes: xlabel='Gender'>


```
In [ ]:
```

In [18]:
```python
my_data1.plot.scatter("Age","Total")
```

Out[18]: <Axes: xlabel='Age', ylabel='Total'>

In [ ]:

In [19]:
```python
my_data1.corr(numeric_only=True)
```

Out[19]:

	Age	Quantitative	Verbal	Technical	Interview	Total	Results
Age	1.000000	-0.473600	-0.406232	-0.746494	0.357157	-0.459172	-0.376011
Quantitative	-0.473600	1.000000	0.363724	0.156132	0.111965	0.760847	0.692700
Verbal	-0.406232	0.363724	1.000000	0.173896	-0.097736	0.574832	0.448660
Technical	-0.746494	0.156132	0.173896	1.000000	-0.013119	0.446550	0.046445
Interview	0.357157	0.111965	-0.097736	-0.013119	1.000000	0.499304	0.500130
Total	-0.459172	0.760847	0.574832	0.446550	0.499304	1.000000	0.783523
Results	-0.376011	0.692700	0.448660	0.046445	0.500130	0.783523	1.000000

In [ ]:

In [20]:
```python
#Find Correlation between two Continuous Variables
my_data1['Age'].corr(my_data1['Interview'])
```

Out[20]: 0.35715686640193034

In [ ]:

In [21]:
```python
my_data1.groupby("Gender")["Total"].mean()
```
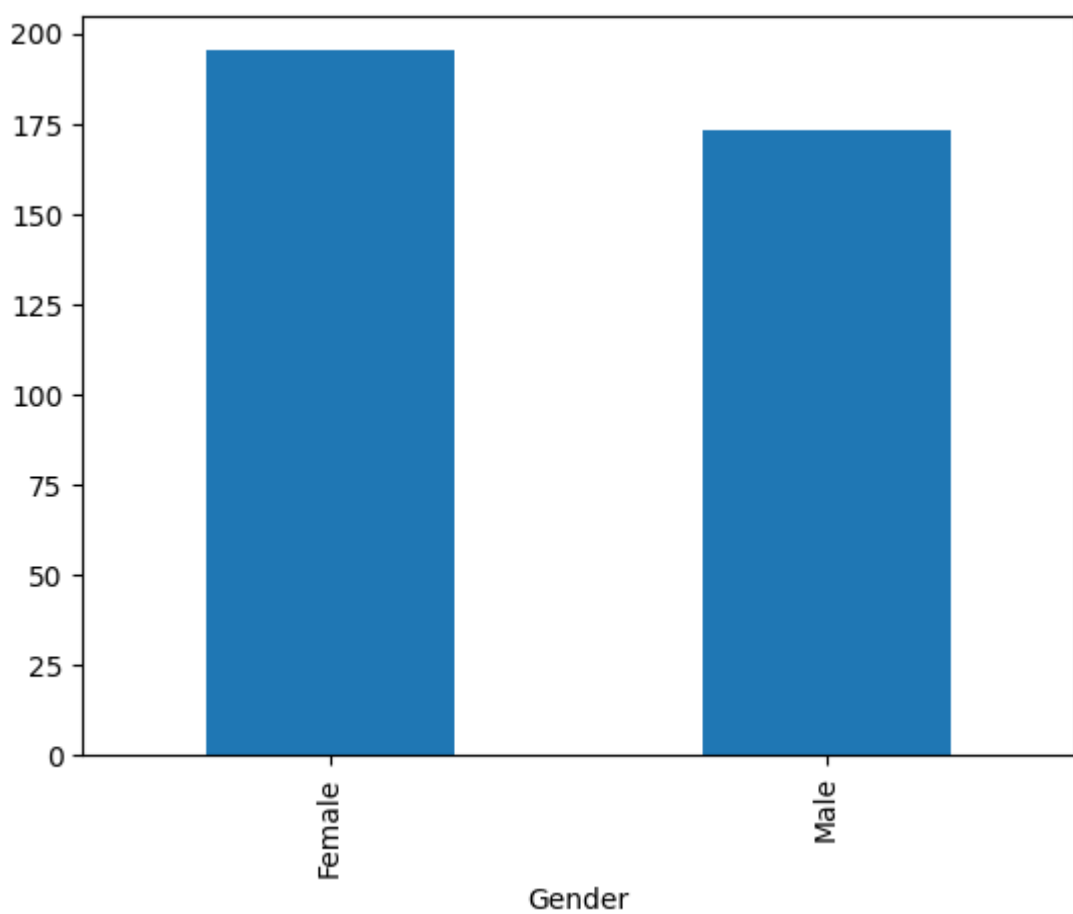
Out[21]:
```
Gender
Female      195.2
Male        173.5
Name: Total, dtype: float64
```

In [ ]:

In [22]:
```python
my_data1.groupby("Gender")["Total"].mean().plot.bar()
```

Out[22]: <Axes: xlabel='Gender'>

In [ ]:

In [23]: `pd.crosstab(my_data1['Gender'],my_data1["Results"])`

Out[23]:

Results	0	1
Gender		
Female	2	3
Male	5	1

In [ ]:

In [ ]: