```
import pandas as pd
import numpy as np

data = {
    "Name": ["Parth", "Avi", "Tejas", "Snehal", "Pankaj", "Ananya"],
    "Gender": ["Male", "Male", "Male", "Female", "Male", "Female"],
    "Marks": [85, 80, 78, 'Nan', 76, 82],
    "Age": [20,21,22,23,24,25]
}
df = pd.DataFrame(data)
print(df)
```

```
     Name  Gender Marks  Age
0   Parth    Male    85   20
1     Avi    Male    80   21
2   Tejas    Male    78   22
3  Snehal  Female   Nan   23
4  Pankaj    Male    76   24
5  Ananya  Female    82   25
```

```
cat=[]
con=[]
for i in df.columns:
    if(df[i].dtypes=="object"):
        cat.append(i)
else:
    con.append(i)
df
```

```
     Name  Gender Marks  Age
0   Parth    Male    85   20
1     Avi    Male    80   21
2   Tejas    Male    78   22
3  Snehal  Female   Nan   23
4  Pankaj    Male    76   24
5  Ananya  Female    82   25
```

```
cat
```

```
['Name', 'Gender', 'Marks']
```

```
con
```

```
['Age']
```

```
c=avg=sum=0
for ele in df['Marks']:
    if str(ele).isnumeric():
        c+=1
        sum+=ele
if c>0:
```

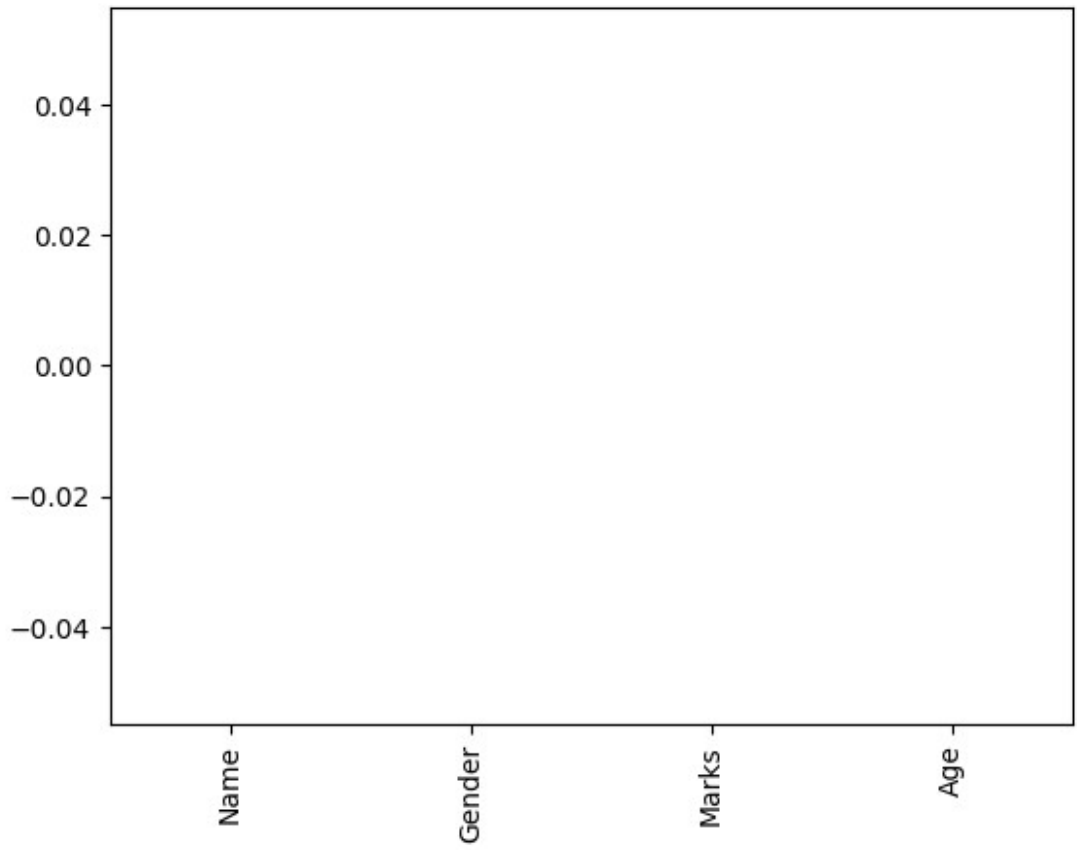
```
    avg=sum/c
df=df.replace(to_replace='Nan', value=avg)
df
```

```
C:\Users\PARTH\AppData\Local\Temp\ipykernel_9488\414823799.py:8:
FutureWarning: Downcasting behavior in `replace` is deprecated and
will be removed in a future version. To retain the old behavior,
explicitly call `result.infer_objects(copy=False)`. To opt-in to the
future behavior, set `pd.set_option('future.no_silent_downcasting',
True)`
  df=df.replace(to_replace='Nan', value=avg)
```

	Name	Gender	Marks	Age
0	Parth	Male	85.0	20
1	Avi	Male	80.0	21
2	Tejas	Male	78.0	22
3	Snehal	Female	80.2	23
4	Pankaj	Male	76.0	24
5	Ananya	Female	82.0	25

```
df.isna().sum().plot(kind="bar")
```

```
<Axes: >
```



```
df['Gender']=df['Gender'].map({'Male':0,'Female':1,}).astype(int)
df
```

```
     Name  Gender  Marks  Age
0   Parth       0   85.0   20
1     Avi       0   80.0   21
2   Tejas       0   78.0   22
3  Snehal       1   80.2   23
4  Pankaj       0   76.0   24
5  Ananya       1   82.0   25
```

```
df=df[df['Marks']>80]
df
```

```
     Name  Gender  Marks  Age
0   Parth       0   85.0   20
3  Snehal       1   80.2   23
5  Ananya       1   82.0   25
```

```
df=df.drop(['Age'], axis=1)
df
```

```
     Name  Gender  Marks
0   Parth       0   85.0
3  Snehal       1   80.2
5  Ananya       1   82.0
```

```
data1 = {
    "Name": ["Parth", "Avi", "Tejas", "Snehal", "Pankaj", "Ananya"],
    "Gender": ["Male", "Male", "Male", "Female", "Male", "Female"],
    "Marks": [85, 80, 78, 'Nan', 76, 82],
    "id": [10,11,12,13,14,15]
}
df1 = pd.DataFrame(data1)
print(df1)
```

```
     Name  Gender Marks  id
0   Parth    Male    85  10
1     Avi    Male    80  11
2   Tejas    Male    78  12
3  Snehal  Female   Nan  13
4  Pankaj    Male    76  14
5  Ananya  Female    82  15
```

```
data2 = {
    "Fee":[50000, 9000, 47000, 10000, 70000,30000],
       "id": [10,11,12,13,14,15]
}
df2 = pd.DataFrame(data2)
print(df)
```

```
     Name  Gender  Marks
0   Parth       0   85.0
3  Snehal       1   80.2
5  Ananya       1   82.0
```

```
df3 = pd.merge(df1, df2)
df3
```

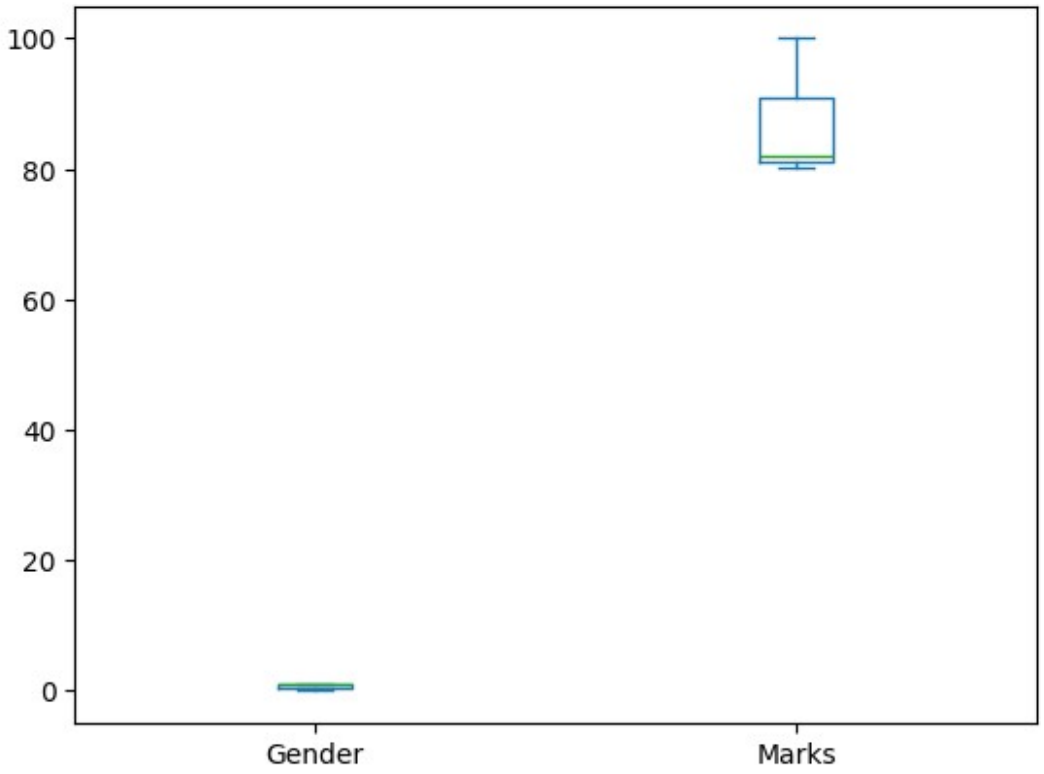
```
     Name  Gender Marks  id    Fee
0   Parth    Male    85  10  50000
1     Avi    Male    80  11   9000
2   Tejas    Male    78  12  47000
3  Snehal  Female   Nan  13  10000
4  Pankaj    Male    76  14  70000
5  Ananya  Female    82  15  30000
```

```
df.loc[0, 'Marks']= 100
print(df)
```

```
     Name  Gender  Marks
0   Parth       0  100.0
3  Snehal       1   80.2
5  Ananya       1   82.0
```

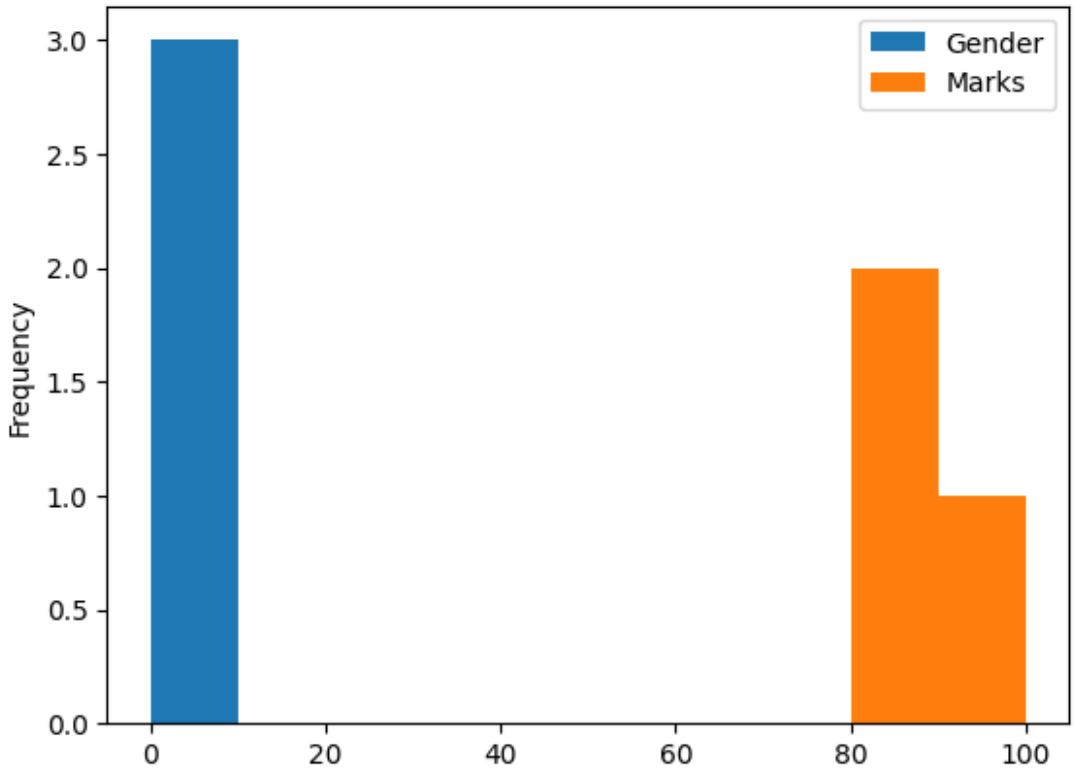
```
df.plot.box()
```

```
<Axes: >
```



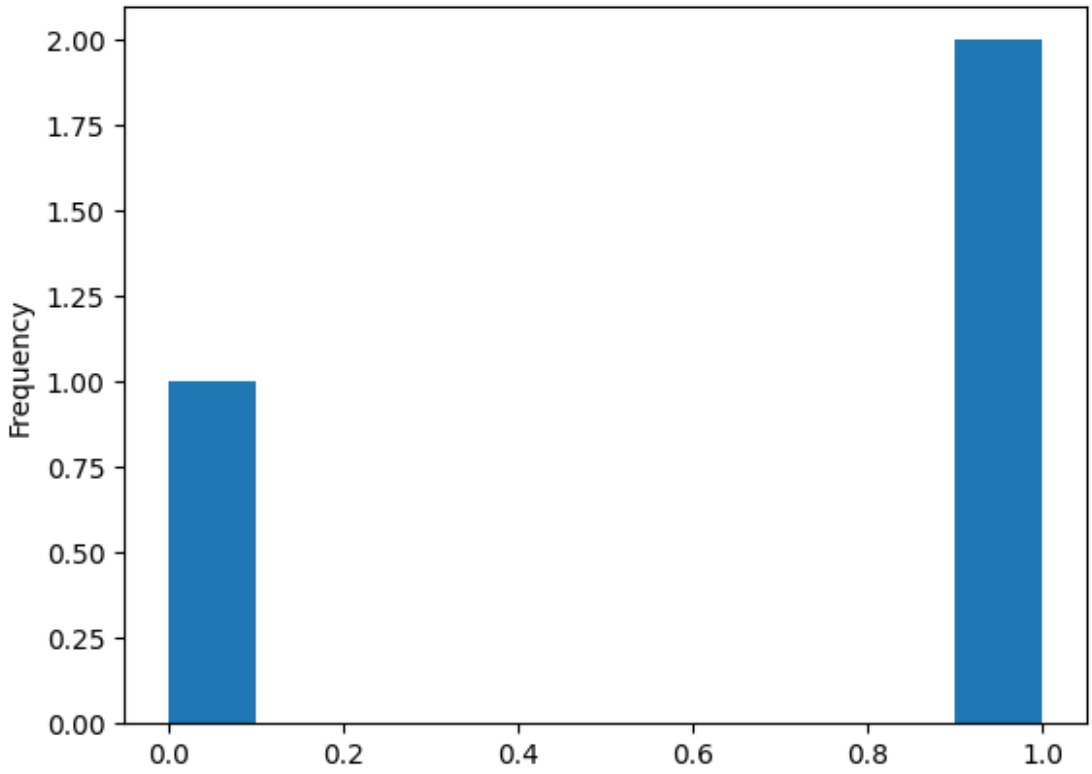

```
df.plot.hist()
```

```
<Axes: ylabel='Frequency'>
```




```
df['Gender'].plot.hist()
```

```
<Axes: ylabel='Frequency'>
```


Frequency
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
0.0
0.2
0.4
0.6
0.8
1.0