In [1]:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as se
```

In [2]:

```python
ap = pd.read_csv("/home/student/Desktop/Academic_Performance.csv")
```

In [3]:

```python
ap.head(6)
```

Out[3]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	80.0	81.0	74	70.0	2020	Male
1	NaN	82.0	87	NaN	2021	Male
2	82.0	86.0	97	80.0	2018	Female
3	85.0	NaN	81	82.0	2019	Male
4	70.0	87.0	80	84.0	2021	Female

In [4]:

```python
ap.isnull()
```

Out[4]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	False	False	False	False	False	False
1	True	False	False	True	False	False
2	False	False	False	False	False	False
3	False	True	False	False	False	False
4	False	False	False	False	False	False

In [6]:

```python
series=pd.isnull(ap["Reading Score"])
ap[series]
```

Out[6]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
3	85.0	NaN	81	82.0	2019	Male

In [7]:

```python
series=pd.isnull(ap["Placement Score"])
ap[series]
```

Out[7]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
1	NaN	82.0	87	NaN	2021	Male

In [8]:

```python
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

In [9]:

```python
ap['Gender'] = le.fit_transform(ap['Gender'])
newdf=ap
ap
```

Out[9]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	80.0	81.0	74	70.0	2020	1
1	NaN	82.0	87	NaN	2021	1
2	82.0	86.0	97	80.0	2018	0
3	85.0	NaN	81	82.0	2019	1
4	70.0	87.0	80	84.0	2021	0

In [10]:

```
ap.dropna(how = 'all')
```

Out[10]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	80.0	81.0	74	70.0	2020	1
1	NaN	82.0	87	NaN	2021	1
2	82.0	86.0	97	80.0	2018	0
3	85.0	NaN	81	82.0	2019	1
4	70.0	87.0	80	84.0	2021	0

In [11]:

```
ap.dropna(axis = 1)
```

Out[11]:

	Writing Score	Club Join Year	Gender
0	74	2020	1
1	87	2021	1
2	97	2018	0
3	81	2019	1
4	80	2021	0

In [12]:

```
new_data = ap.dropna(axis = 0,how='any')
new_data
```

Out[12]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	80.0	81.0	74	70.0	2020	1
2	82.0	86.0	97	80.0	2018	0
4	70.0	87.0	80	84.0	2021	0

In [13]:

```
print(np.where(ap['Reading Score']<25))
print(np.where(ap['Writing Score']<30))
```

```
(array([], dtype=int64),)
(array([], dtype=int64),)
```

In [19]:

```
sorted_rscore = sorted(ap['Writing Score'])
```

In [20]:

```
q1 = np.percentile(sorted_rscore,25)
q3 = np.percentile(sorted_rscore,75)
print(q1,q3)
```

```
80.0 87.0
```

In [21]:

```
IQR = q1-q3
```

In [22]:

```
lwr_bound = q1-(1.5*IQR)
upr_bound = q1+(1.5*IQR)
print(lwr_bound,upr_bound)
```

```
90.5 69.5
```

In [24]:

```
r_outliers = []
for i in sorted_rscore:
    if(i<lwr_bound or i>upr_bound):
        r_outliers.append(i)
print(r_outliers)
```

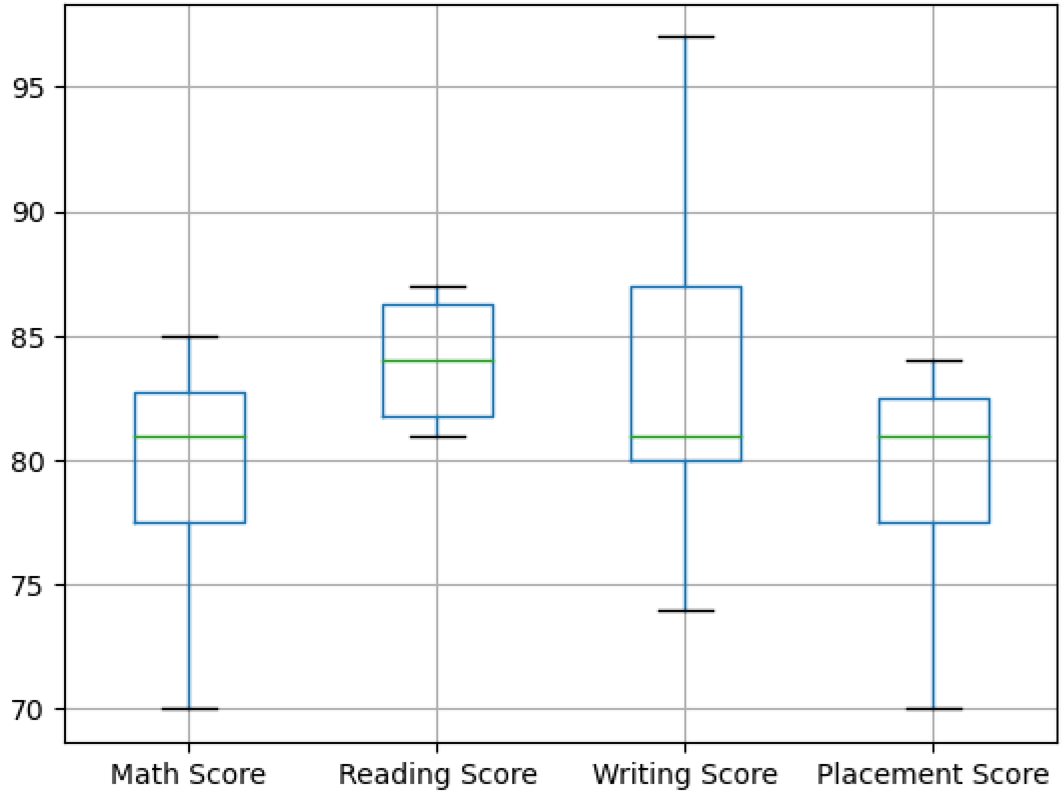
```
[74, 80, 81, 87, 97]
```

In [25]:

```
col = ['Math Score','Reading Score' , 'Writing Score','Placement Score']
ap.boxplot(col)
```

Out[25]:

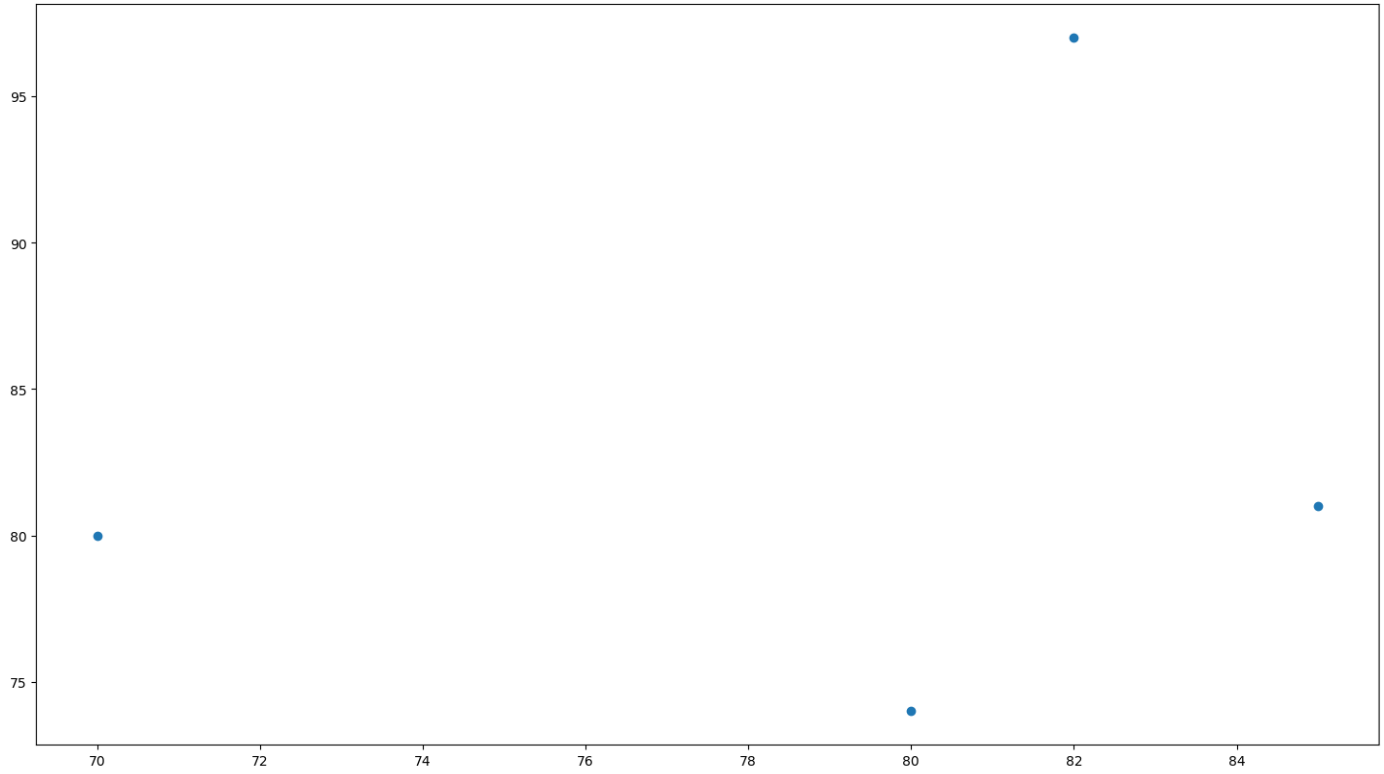
```
<Axes: >
```



In [29]:

```
fig,ax = plt.subplots(figsize = (18,10))
ax.scatter(ap['Math Score'],ap['Writing Score'])
plt.show()
```

Loading [MathJax]/extensions/Safe.js



In [30]:

```
import numpy as np
from scipy import stats
```

In [32]:

```
z = np.abs(stats.zscore(ap['Writing Score']))
print(z)
```

```
0    1.259311
1    0.411204
2    1.696215
3    0.359803
4    0.488304
Name: Writing Score, dtype: float64
```

In [33]:

```
threshold = 0.18
```

In [34]:

```
sample_outliers = np.where(z <threshold)
sample_outliers
```

Out[34]:

```
(array([], dtype=int64),)
```

In [35]:

```
sorted_rscore =sorted(ap['Reading Score'])
```

In [36]:

```
sorted_rscore
```

Out[36]:

```
[81.0, 82.0, 86.0, nan, 87.0]
```

In [38]:

```
new_df=ap
for i in sample_outliers:
    new_df.drop(i,inplace=True)
new_df
```

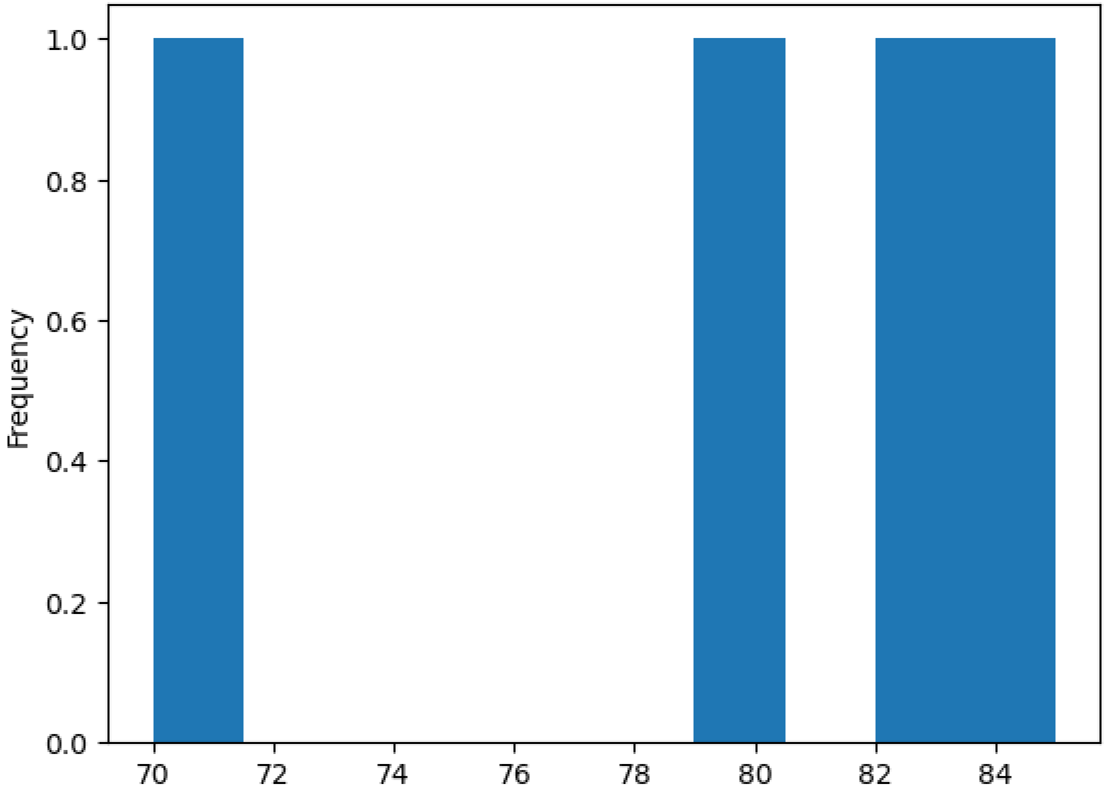
Loading [MathJax]/extensions/Safe.js

Out[38]:

	Math Score	Reading Score	Writing Score	Placement Score	Club Join Year	Gender
0	80.0	81.0	74	70.0	2020	1
1	NaN	82.0	87	NaN	2021	1
2	82.0	86.0	97	80.0	2018	0
3	85.0	NaN	81	82.0	2019	1
4	70.0	87.0	80	84.0	2021	0

In [41]:

```
ap['Math Score'].plot(kind = 'hist')
plt.show()
```



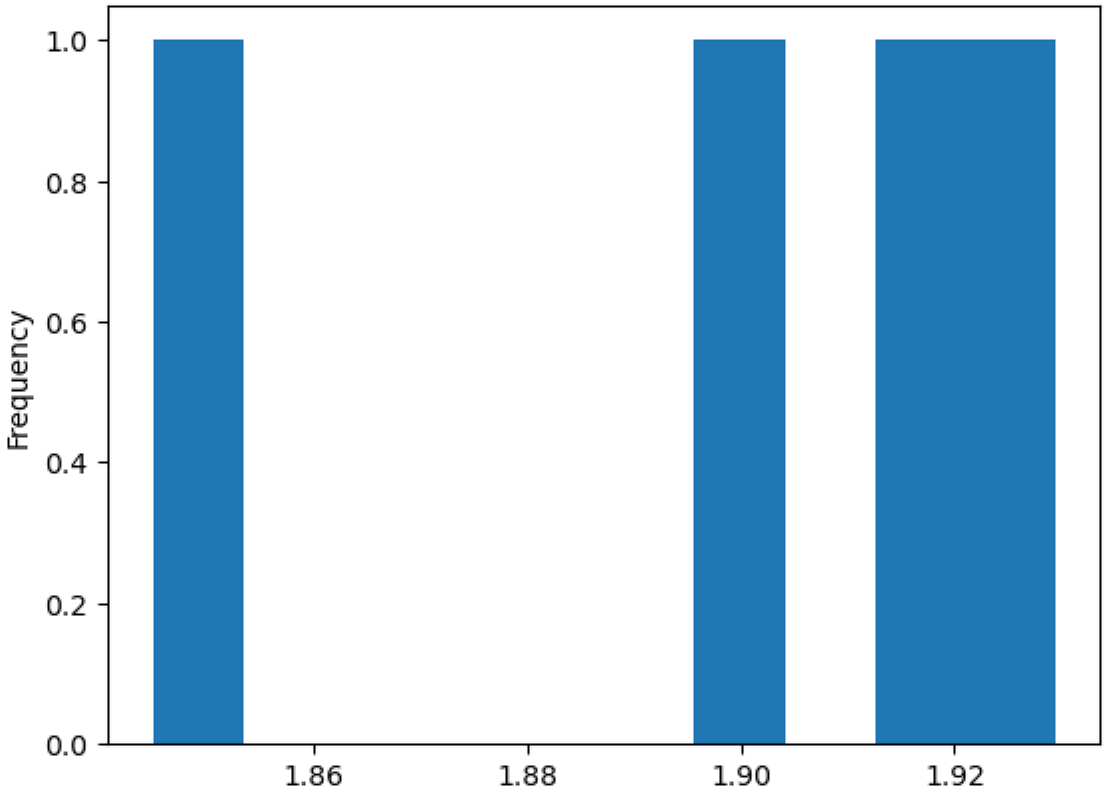
In [45]:

```
ap['log_math'] = np.log10(ap['Math Score'])
```

In [46]:

```
ap['log_math'].plot(kind = 'hist')
plt.show()
```

Loading [MathJax]/extensions/Safe.js



```
In [ ]:
```

Loading [MathJax]/extensions/Safe.js