In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as se
```

In [2]:
```
sp=pd.read_csv("/home/student/Desktop/academicperformance.csv")
```

In [3]:
```
sp.head(6)
```

Out[3]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	male
1	81	NaN	87	80.0	2021	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	NaN	2019	male
4	70	87.0	80	84.0	2021	female

In [4]:
```
sp.isnull()
```

Out[4]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	False	False	False	False	False	False
1	False	True	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	True	False	False
4	False	False	False	False	False	False

In [5]:
```
series=pd.isnull(sp["Math score"])
sp[series]
```

Out[5]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender

In [6]:
```
series=pd.isnull(sp["Reading score"])
sp[series]
```

Out[6]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
1	81	NaN	87	80.0	2021	male

In [11]:
```
sp.notnull()
```

Out[11]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	True	True	True	True	True	True
1	True	False	True	True	True	True
2	True	True	True	True	True	True
3	True	True	True	False	True	True
4	True	True	True	True	True	True

In [15]:
```
series1=pd.notnull(sp["Reading score"])
sp[series1]
```

Out[15]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	NaN	2019	male
4	70	87.0	80	84.0	2021	female

In [17]:
```
sp.fillna(12)
```

Out[17]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	male
1	81	12.0	87	80.0	2021	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	12.0	2019	male
4	70	87.0	80	84.0	2021	female

In [38]:
```
sp.dropna(axis = 1)
```

Out[38]:

	Math score	Writing score	Club Join year	Gender
0	80	74	2020	1
1	81	87	2021	1
2	82	97	2018	0
3	83	81	2019	1
4	70	80	2021	0

In [41]:
```
new_data=sp.dropna(axis = 0,how='any')
new_data
```

Out[41]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	1
2	82	86.0	97	82.0	2018	0
4	70	87.0	80	84.0	2021	0

In [29]:
```
sp=pd.read_csv("/home/student/Desktop/academicperformance.csv")
```

In [30]:
```
sp.head(6)
```

Out[30]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	male
1	81	NaN	87	80.0	2021	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	NaN	2019	male
4	70	87.0	80	84.0	2021	female

In [31]:
```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
```

In [32]:
```
sp['Gender']=le.fit_transform(sp['Gender'])
newdf=sp
sp
```

Out[32]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	1
1	81	NaN	87	80.0	2021	1
2	82	86.0	97	82.0	2018	0
3	83	85.0	81	NaN	2019	1
4	70	87.0	80	84.0	2021	0

In [42]:
```
sp.dropna(how='all')
```

Out[42]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	1
1	81	NaN	87	80.0	2021	1
2	82	86.0	97	82.0	2018	0
3	83	85.0	81	NaN	2019	1
4	70	87.0	80	84.0	2021	0

In [53]:
```
print(np.where(sp['Math score']<82))
print(np.where(sp['Writing score']<80))
```

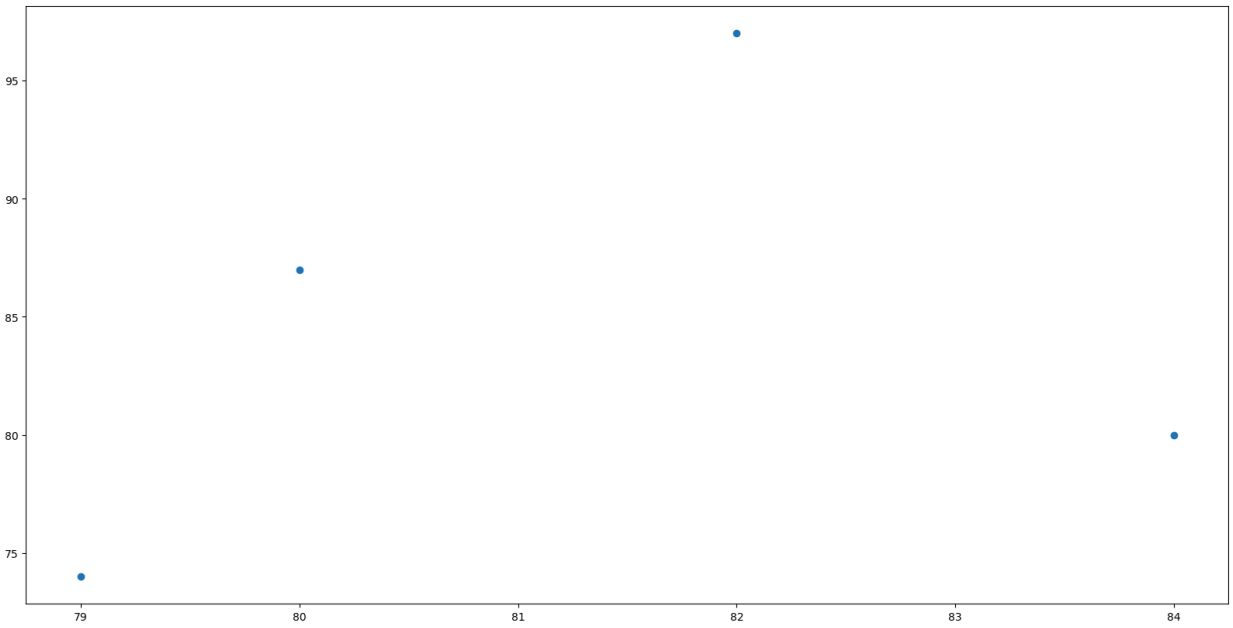
```
(array([0, 1, 4]),)
(array([0]),)
```

In [60]:
```
sorted_rscore=sorted(sp['Math score'])
```

In [61]:
```
q1=np.percentile(sorted_rscore,82)
q3=np.percentile(sorted_rscore,70)
print(q1,q3)
```

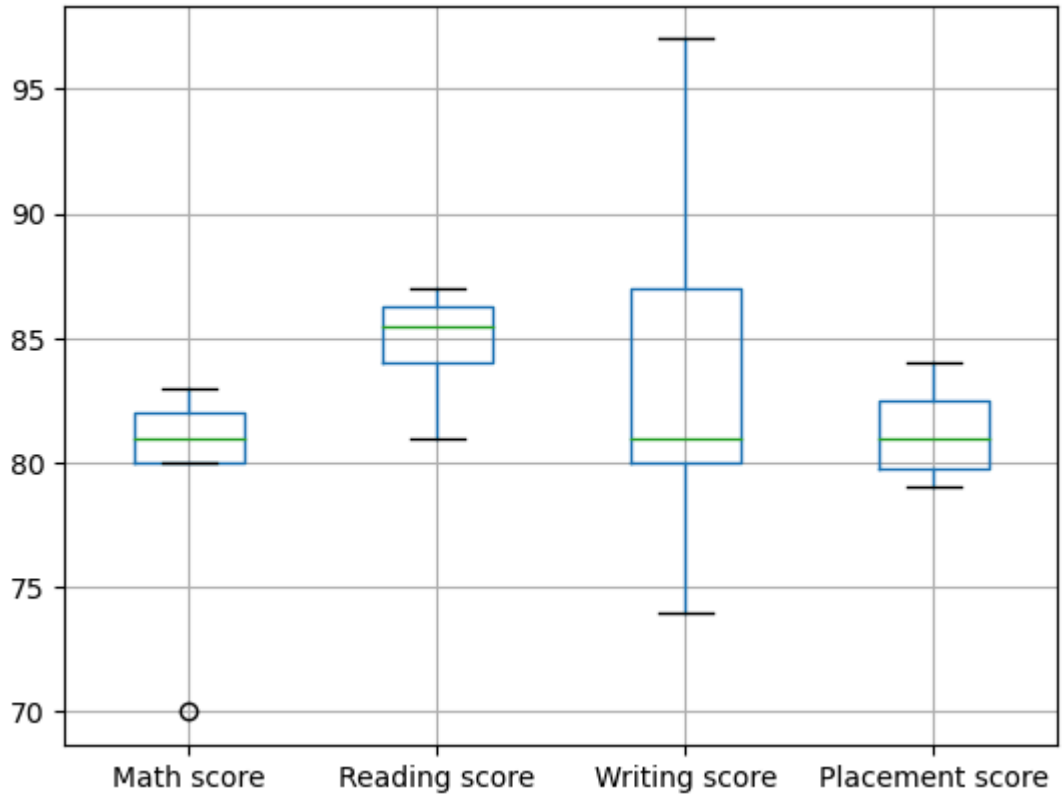
```
82.28 81.8
```

In [76]:
```
fig,ax=plt.subplots(figsize =(20,10))
ax.scatter(sp['Placement score'],sp['Writing score'])
plt.show()
```



In [65]:
```
col=['Math score','Reading score','Writing score','Placement score']
sp.boxplot(col)
```

Out[65]:
```
<Axes: >
```



In [77]:
```
import numpy as np
from scipy import stats
```

In [78]:
```
z = np.abs(stats.zscore(sp['Math score']))
print(z)
```

```
0    0.169944
1    0.382373
2    0.594803
3    0.807233
4    1.954353
Name: Math score, dtype: float64
```

In [83]:
```
threshold =0.20
```

In [85]:
```
sample_outliers = np.where(z<threshold)
sample_outliers
```

Out[85]:
```
(array([0]),)
```

In [86]:
```
sorted_rscore=sorted(sp['Math score'])
```

In [87]:
```
sorted_rscore
```

Out[87]:
```
[70, 80, 81, 82, 83]
```

In [96]:
```
sp=pd.read_csv("/home/student/Desktop/academicperformance.csv")
```

In [97]:
```
new_df=sp
for i in sample_outliers:

```

```
    new_df.drop(i,inplace=True)
new_df
```

Out[97]:

	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
1	81	NaN	87	80.0	2021	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	NaN	2019	male
4	70	87.0	80	84.0	2021	female

In [98]:
```
import matplotlib.pyplot as plt
```

In [99]:
```
import pandas as pd
sp=pd.read_csv("/home/student/Desktop/academicperformance.csv")
```

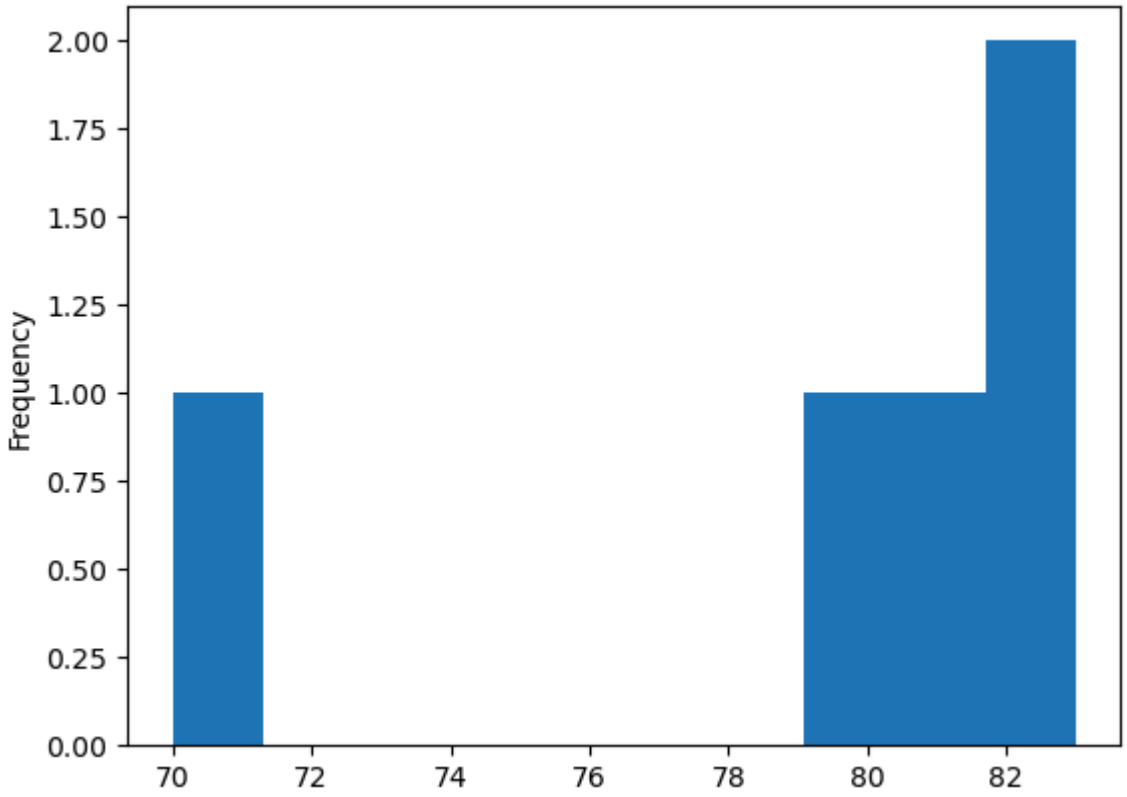
In [100…
```
df=pd.read_csv("/home/student/Desktop/academicperformance.csv")
```

In [101…
```
df
```

Out[101]:

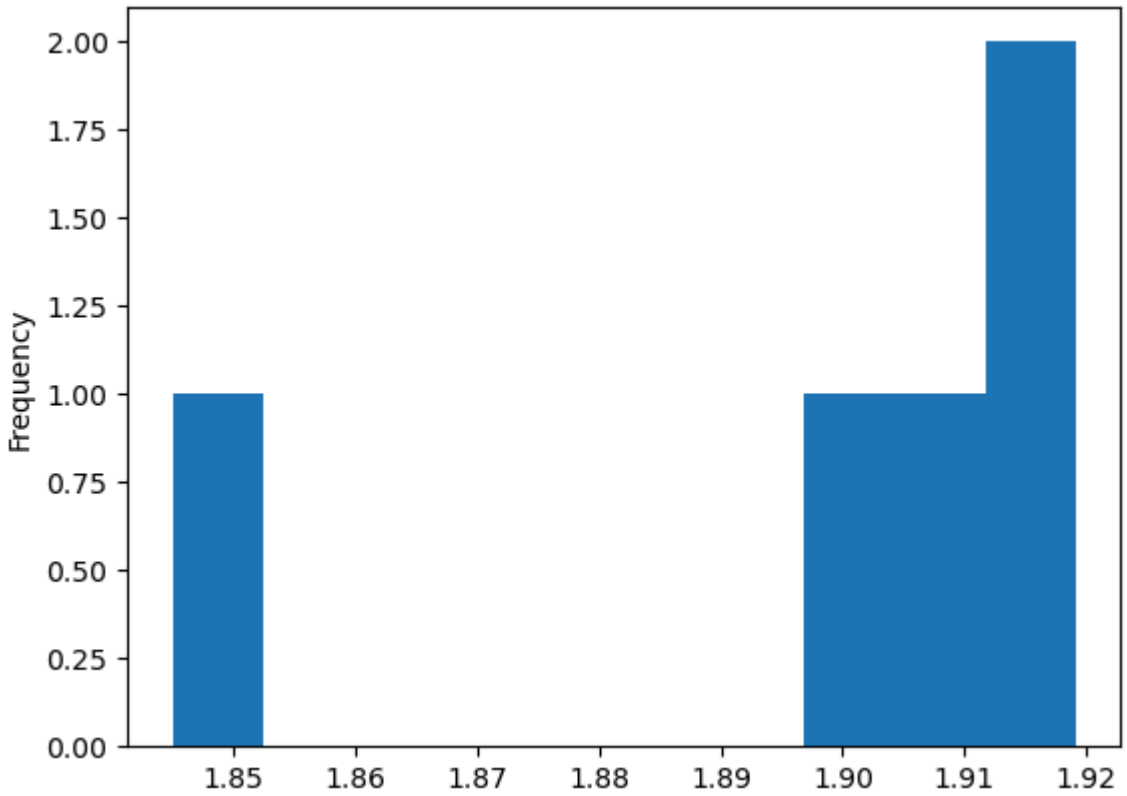
	Math score	Reading score	Writing score	Placement score	Club Join year	Gender
0	80	81.0	74	79.0	2020	male
1	81	NaN	87	80.0	2021	male
2	82	86.0	97	82.0	2018	female
3	83	85.0	81	NaN	2019	male
4	70	87.0	80	84.0	2021	female

In [102…
```
df['Math score'].plot(kind = 'hist')
plt.show()
```



In [103…
```
import numpy as np
df['log_math']=np.log10(df['Math score'])
```

In [104…
```
df['log_math'].plot(kind = 'hist')
plt.show()
```



In [ ]: