In [24]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
df=pd.read_csv("StudentsPerformance.csv")
df.isnull()
```

Out[24]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	True	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	True	False	False	False
8	False	False	False	False	False
9	False	False	False	False	False
10	True	False	False	False	False
11	False	False	False	False	False
12	False	False	False	False	False
13	False	False	True	False	False
14	False	False	False	False	True
15	False	False	False	False	False
16	False	False	False	False	False
17	False	False	False	False	False
18	False	False	False	False	False
19	False	False	False	False	False
20	True	True	False	False	False
21	False	False	False	False	False
22	False	False	False	False	False
23	False	False	False	False	False
24	False	False	False	False	False
25	False	False	False	False	False
26	False	True	False	False	False
27	False	False	False	False	False
28	False	False	False	False	False
29	False	False	False	False	False

In [8]:
```
series = pd.isnull(df["Math_Score"])
df[series]
```

Out[8]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
10	NaN	74.0	74.0	88.0	2019.0
20	NaN	NaN	79.0	83.0	2019.0

In [9]:
```
df.notnull()
```

Out[9]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	True	True	True	True	True
1	True	True	True	True	True
2	True	True	True	False	True
3	True	True	True	True	True
4	True	True	True	True	True
5	True	True	True	True	True
6	True	True	True	True	True
7	True	False	True	True	True
8	True	True	True	True	True
9	True	True	True	True	True
10	False	True	True	True	True
11	True	True	True	True	True
12	True	True	True	True	True
13	True	True	False	True	True
14	True	True	True	True	False
15	True	True	True	True	True
16	True	True	True	True	True
17	True	True	True	True	True
18	True	True	True	True	True
19	True	True	True	True	True
20	False	False	True	True	True
21	True	True	True	True	True
22	True	True	True	True	True
23	True	True	True	True	True
24	True	True	True	True	True
25	True	True	True	True	True
26	True	False	True	True	True
27	True	True	True	True	True
28	True	True	True	True	True
29	True	True	True	True	True

In [10]:
```
series = pd.notnull(df["Math_Score"])
df[series]
```

Out[10]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	74.0	75.0	80.0	89.0	2020.0
1	75.0	77.0	74.0	84.0	2018.0
2	77.0	78.0	78.0	NaN	2019.0
3	68.0	79.0	66.0	81.0	2020.0
4	80.0	83.0	77.0	85.0	2021.0
5	61.0	84.0	64.0	76.0	2021.0
6	73.0	74.0	74.0	79.0	2019.0
7	68.0	NaN	80.0	87.0	2020.0
8	63.0	78.0	78.0	75.0	2018.0
9	63.0	79.0	79.0	86.0	2019.0
11	71.0	75.0	75.0	88.0	2021.0
12	65.0	76.0	60.0	89.0	2020.0
13	75.0	77.0	NaN	81.0	2019.0
14	64.0	81.0	68.0	80.0	NaN
15	77.0	82.0	68.0	81.0	2020.0
16	78.0	83.0	70.0	89.0	2018.0
17	79.0	84.0	74.0	86.0	2021.0
18	63.0	80.0	76.0	80.0	2019.0
19	73.0	80.0	78.0	85.0	2019.0
21	70.0	80.0	80.0	82.0	2018.0
22	74.0	76.0	79.0	80.0	2021.0
23	63.0	78.0	74.0	80.0	2021.0
24	69.0	76.0	75.0	86.0	2020.0
25	68.0	79.0	60.0	75.0	2020.0
26	72.0	NaN	64.0	82.0	2019.0
27	65.0	80.0	68.0	79.0	2018.0
28	69.0	78.0	68.0	80.0	2020.0
29	79.0	74.0	74.0	86.0	2021.0

In [11]:
```
ndf=df
ndf.fillna(0)
```

Out[11]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	74.0	75.0	80.0	89.0	2020.0
1	75.0	77.0	74.0	84.0	2018.0
2	77.0	78.0	78.0	0.0	2019.0
3	68.0	79.0	66.0	81.0	2020.0
4	80.0	83.0	77.0	85.0	2021.0
5	61.0	84.0	64.0	76.0	2021.0
6	73.0	74.0	74.0	79.0	2019.0
7	68.0	0.0	80.0	87.0	2020.0
8	63.0	78.0	78.0	75.0	2018.0
9	63.0	79.0	79.0	86.0	2019.0
10	0.0	74.0	74.0	88.0	2019.0
11	71.0	75.0	75.0	88.0	2021.0
12	65.0	76.0	60.0	89.0	2020.0
13	75.0	77.0	0.0	81.0	2019.0
14	64.0	81.0	68.0	80.0	0.0
15	77.0	82.0	68.0	81.0	2020.0
16	78.0	83.0	70.0	89.0	2018.0
17	79.0	84.0	74.0	86.0	2021.0
18	63.0	80.0	76.0	80.0	2019.0
19	73.0	80.0	78.0	85.0	2019.0
20	0.0	0.0	79.0	83.0	2019.0
21	70.0	80.0	80.0	82.0	2018.0
22	74.0	76.0	79.0	80.0	2021.0
23	63.0	78.0	74.0	80.0	2021.0
24	69.0	76.0	75.0	86.0	2020.0
25	68.0	79.0	60.0	75.0	2020.0
26	72.0	0.0	64.0	82.0	2019.0
27	65.0	80.0	68.0	79.0	2018.0
28	69.0	78.0	68.0	80.0	2020.0
29	79.0	74.0	74.0	86.0	2021.0

In [2]:
```
m_v=df['Math_Score'].mean()
df['Math_Score'].fillna(value=m_v, inplace=True)
df
```

Out[2]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	74.000000	75.0	80.0	89.0	2020.0
1	75.000000	77.0	74.0	84.0	2018.0
2	77.000000	78.0	78.0	NaN	2019.0
3	68.000000	79.0	66.0	81.0	2020.0
4	80.000000	83.0	77.0	85.0	2021.0
5	61.000000	84.0	64.0	76.0	2021.0
6	73.000000	74.0	74.0	79.0	2019.0
7	68.000000	NaN	80.0	87.0	2020.0
8	63.000000	78.0	78.0	75.0	2018.0
9	63.000000	79.0	79.0	86.0	2019.0
10	70.571429	74.0	74.0	88.0	2019.0
11	71.000000	75.0	75.0	88.0	2021.0
12	65.000000	76.0	60.0	89.0	2020.0
13	75.000000	77.0	NaN	81.0	2019.0
14	64.000000	81.0	68.0	80.0	NaN
15	77.000000	82.0	68.0	81.0	2020.0
16	78.000000	83.0	70.0	89.0	2018.0
17	79.000000	84.0	74.0	86.0	2021.0
18	63.000000	80.0	76.0	80.0	2019.0
19	73.000000	80.0	78.0	85.0	2019.0
20	70.571429	NaN	79.0	83.0	2019.0
21	70.000000	80.0	80.0	82.0	2018.0
22	74.000000	76.0	79.0	80.0	2021.0
23	63.000000	78.0	74.0	80.0	2021.0
24	69.000000	76.0	75.0	86.0	2020.0
25	68.000000	79.0	60.0	75.0	2020.0
26	72.000000	NaN	64.0	82.0	2019.0
27	65.000000	80.0	68.0	79.0	2018.0
28	69.000000	78.0	68.0	80.0	2020.0
29	79.000000	74.0	74.0	86.0	2021.0

In [4]:
```
df.replace(to_replace = np.nan, value = -99)
```

Out[4]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	74.000000	75.0	80.0	89.0	2020.0
1	75.000000	77.0	74.0	84.0	2018.0
2	77.000000	78.0	78.0	-99.0	2019.0
3	68.000000	79.0	66.0	81.0	2020.0
4	80.000000	83.0	77.0	85.0	2021.0
5	61.000000	84.0	64.0	76.0	2021.0
6	73.000000	74.0	74.0	79.0	2019.0
7	68.000000	-99.0	80.0	87.0	2020.0
8	63.000000	78.0	78.0	75.0	2018.0
9	63.000000	79.0	79.0	86.0	2019.0
10	70.571429	74.0	74.0	88.0	2019.0
11	71.000000	75.0	75.0	88.0	2021.0
12	65.000000	76.0	60.0	89.0	2020.0
13	75.000000	77.0	-99.0	81.0	2019.0
14	64.000000	81.0	68.0	80.0	-99.0
15	77.000000	82.0	68.0	81.0	2020.0
16	78.000000	83.0	70.0	89.0	2018.0
17	79.000000	84.0	74.0	86.0	2021.0
18	63.000000	80.0	76.0	80.0	2019.0
19	73.000000	80.0	78.0	85.0	2019.0
20	70.571429	-99.0	79.0	83.0	2019.0
21	70.000000	80.0	80.0	82.0	2018.0
22	74.000000	76.0	79.0	80.0	2021.0
23	63.000000	78.0	74.0	80.0	2021.0
24	69.000000	76.0	75.0	86.0	2020.0
25	68.000000	79.0	60.0	75.0	2020.0
26	72.000000	-99.0	64.0	82.0	2019.0
27	65.000000	80.0	68.0	79.0	2018.0
28	69.000000	78.0	68.0	80.0	2020.0
29	79.000000	74.0	74.0	86.0	2021.0

In [5]:
```
df.dropna()
```

Out[5]:

	Math_Score	Reading_Score	Writing_Score	Placement_Score	Club_Join_Date .
0	74.000000	75.0	80.0	89.0	2020.0
1	75.000000	77.0	74.0	84.0	2018.0
3	68.000000	79.0	66.0	81.0	2020.0
4	80.000000	83.0	77.0	85.0	2021.0
5	61.000000	84.0	64.0	76.0	2021.0
6	73.000000	74.0	74.0	79.0	2019.0
8	63.000000	78.0	78.0	75.0	2018.0
9	63.000000	79.0	79.0	86.0	2019.0
10	70.571429	74.0	74.0	88.0	2019.0
11	71.000000	75.0	75.0	88.0	2021.0
12	65.000000	76.0	60.0	89.0	2020.0
15	77.000000	82.0	68.0	81.0	2020.0
16	78.000000	83.0	70.0	89.0	2018.0
17	79.000000	84.0	74.0	86.0	2021.0
18	63.000000	80.0	76.0	80.0	2019.0
19	73.000000	80.0	78.0	85.0	2019.0
21	70.000000	80.0	80.0	82.0	2018.0
22	74.000000	76.0	79.0	80.0	2021.0
23	63.000000	78.0	74.0	80.0	2021.0
24	69.000000	76.0	75.0	86.0	2020.0
25	68.000000	79.0	60.0	75.0	2020.0
27	65.000000	80.0	68.0	79.0	2018.0
28	69.000000	78.0	68.0	80.0	2020.0
29	79.000000	74.0	74.0	86.0	2021.0

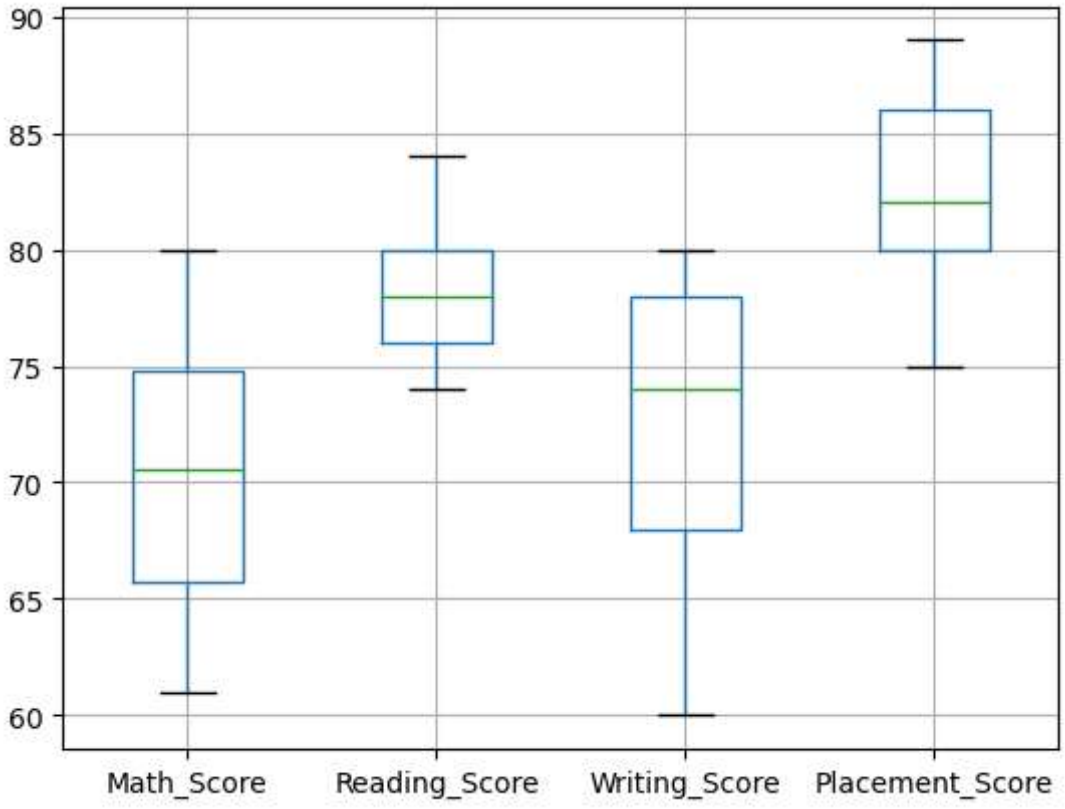
In [6]:
```
df.dropna(axis = 1)
```

Out[6]:

	Math_Score
0	74.000000
1	75.000000
2	77.000000
3	68.000000
4	80.000000
5	61.000000
6	73.000000
7	68.000000
8	63.000000
9	63.000000
10	70.571429
11	71.000000
12	65.000000
13	75.000000
14	64.000000
15	77.000000
16	78.000000
17	79.000000
18	63.000000
19	73.000000
20	70.571429
21	70.000000
22	74.000000
23	63.000000
24	69.000000
25	68.000000
26	72.000000
27	65.000000
28	69.000000
29	79.000000

In [7]:
```
col = ['Math_Score', 'Reading_Score' , 'Writing_Score','Placement_Score']
df.boxplot(col)
```

Out[7]: <Axes: >



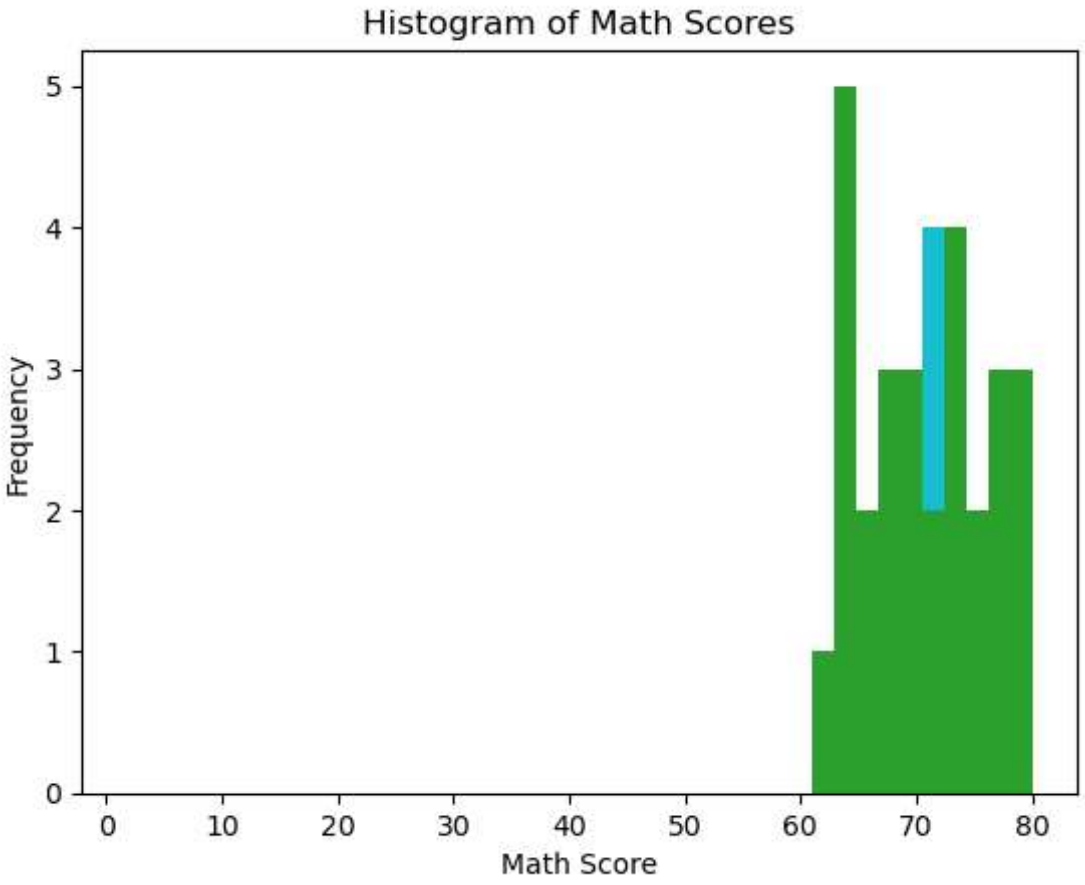
In [11]:
```
print(np.where(df['Math_Score']>75))
```

```
(array([ 2,  4, 15, 16, 17, 29], dtype=int64),)
```

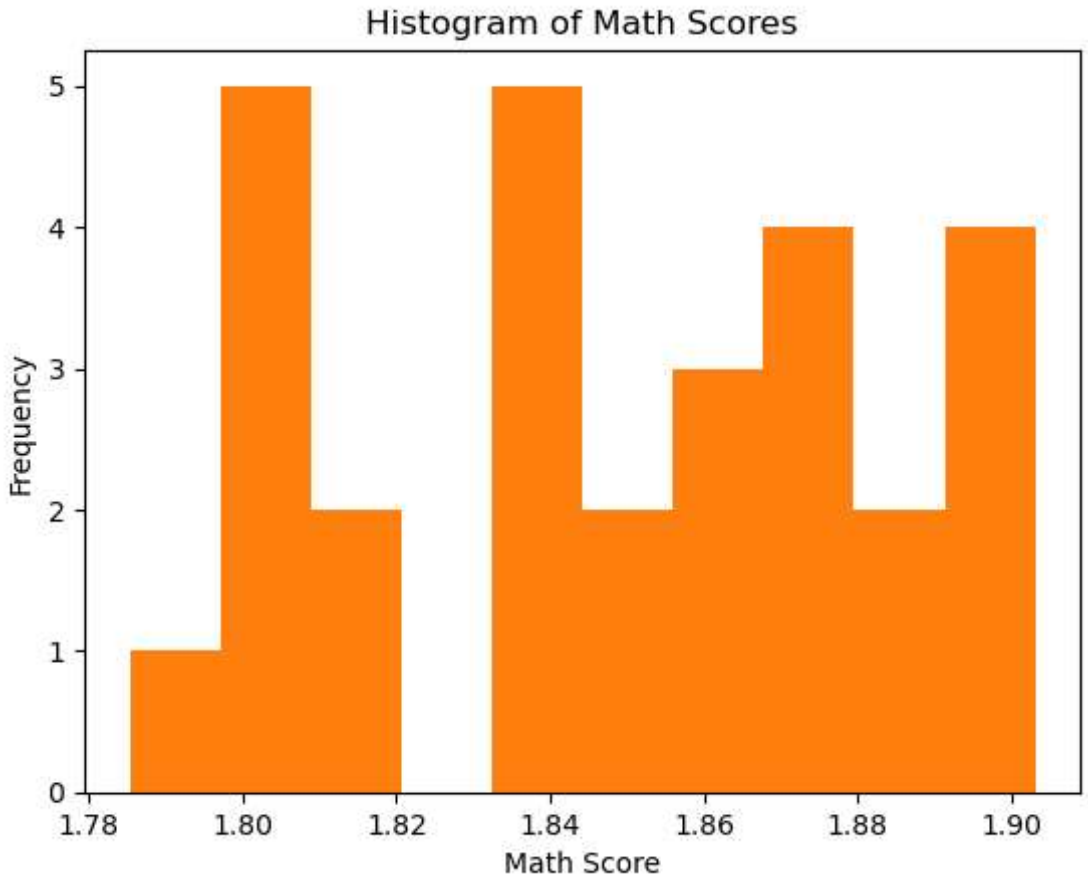
In [12]:
```
print(np.where(df['Reading_Score']<85))
```

```
(array([ 0,  1,  2,  3,  4,  5,  6,  8,  9, 10, 11, 12, 13, 14, 15, 16, 17,
       18, 19, 21, 22, 23, 24, 25, 27, 28, 29], dtype=int64),)
```

In [33]:
```
df['Math_Score'].plot(kind ='hist')
plt.xlabel('Math Score')
plt.ylabel('Frequency')
plt.title('Histogram of Math Scores')
plt.show()
```



In [35]:
```
df['log_math'] = np.log10(df['Math_Score'])
df['log_math'].plot(kind = 'hist')
plt.xlabel('Math Score')
plt.ylabel('Frequency')
plt.title('Histogram of Math Scores')
plt.show()
```



In [ ]: