

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats as st
```

```
In [2]: df = pd.read_csv('StudentsPerformanceTest1.csv')
```

```
In [3]: df
```

```
Out[3]:
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72.0	72.0	74.0	78.0	1	Pune
1	female	69.0	90.0	88.0	NaN	2	NaN
2	female	90.0	95.0	93.0	74.0	2	Nashik
3	male	47.0	57.0	NaN	78.0	1	NaN
4	male	NaN	78.0	75.0	81.0	3	Pune
5	female	71.0	NaN	78.0	70.0	4	NaN
6	male	12.0	44.0	52.0	12.0	2	Nashik
7	male	NaN	65.0	67.0	49.0	1	Pune
8	male	5.0	77.0	89.0	55.0	0	NaN

```
In [4]: df2 = df.copy()
```

```
In [5]: df2
```

```
Out[5]:
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72.0	72.0	74.0	78.0	1	Pune
1	female	69.0	90.0	88.0	NaN	2	NaN
2	female	90.0	95.0	93.0	74.0	2	Nashik
3	male	47.0	57.0	NaN	78.0	1	NaN
4	male	NaN	78.0	75.0	81.0	3	Pune
5	female	71.0	NaN	78.0	70.0	4	NaN
6	male	12.0	44.0	52.0	12.0	2	Nashik
7	male	NaN	65.0	67.0	49.0	1	Pune
8	male	5.0	77.0	89.0	55.0	0	NaN

```
In [6]: mean = df2['math score'].mean()
```

```
In [7]: mean
```

```
Out[7]: 52.285714285714285
```

```
In [8]: median = df2['math score'].median()
```

In [9]:

median

Out[9]:

69.0

In [10]:

mode = df2.mode().values[0]

In [11]:

mode

Out[11]:

array(['male', 5.0, 44.0, 52.0, 78.0, 1.0, 'Pune'], dtype=object)

In [12]:

df2.isnull()

Out[12]:

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	False	False	False	False	False	False	False
1	False	False	False	False	True	False	True
2	False	False	False	False	False	False	False
3	False	False	False	True	False	False	True
4	False	True	False	False	False	False	False
5	False	False	True	False	False	False	True
6	False	False	False	False	False	False	False
7	False	True	False	False	False	False	False
8	False	False	False	False	False	False	True

In [13]:

num_columns = [col for col in df2.columns if (df2[col].dtype == 'float')]

In [14]:

num_columns

Out[14]:

['math score', 'reading score', 'writing score', 'Placement Score']

In [15]:

for i in num_columns:
 df2[i].fillna(round(mean), inplace=True)

In [16]:

df2

Out[16]:

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72.0	72.0	74.0	78.0	1	Pune
1	female	69.0	90.0	88.0	52.0	2	NaN
2	female	90.0	95.0	93.0	74.0	2	Nashik
3	male	47.0	57.0	52.0	78.0	1	NaN
4	male	52.0	78.0	75.0	81.0	3	Pune
5	female	71.0	52.0	78.0	70.0	4	NaN
6	male	12.0	44.0	52.0	12.0	2	Nashik
7	male	52.0	65.0	67.0	49.0	1	Pune
8	male	5.0	77.0	89.0	55.0	0	NaN

In [17]: `print(df2['Region'].dtype)`

object

In [18]: `str_columns = [col for col in df2.columns if (df2[col].dtype=='object')]`In [19]: `str_columns`Out[19]: `['gender', 'Region']`

```
In [20]: for i in str_columns:
          if(i=='gender'):
              continue
          else:
              df2[i].fillna(df2[i].mode().values[0], inplace=True)
```

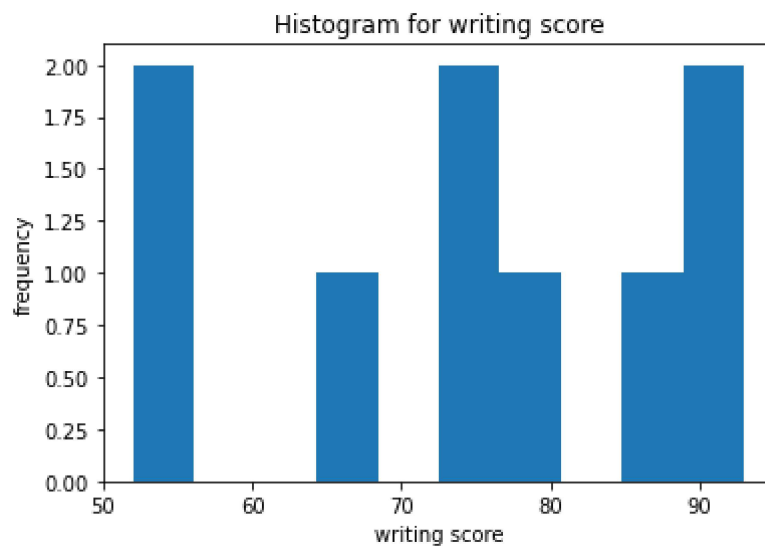
In [21]: `df2`

Out[21]:

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72.0	72.0	74.0	78.0	1	Pune
1	female	69.0	90.0	88.0	52.0	2	Pune
2	female	90.0	95.0	93.0	74.0	2	Nashik
3	male	47.0	57.0	52.0	78.0	1	Pune
4	male	52.0	78.0	75.0	81.0	3	Pune
5	female	71.0	52.0	78.0	70.0	4	Pune
6	male	12.0	44.0	52.0	12.0	2	Nashik
7	male	52.0	65.0	67.0	49.0	1	Pune
8	male	5.0	77.0	89.0	55.0	0	Pune

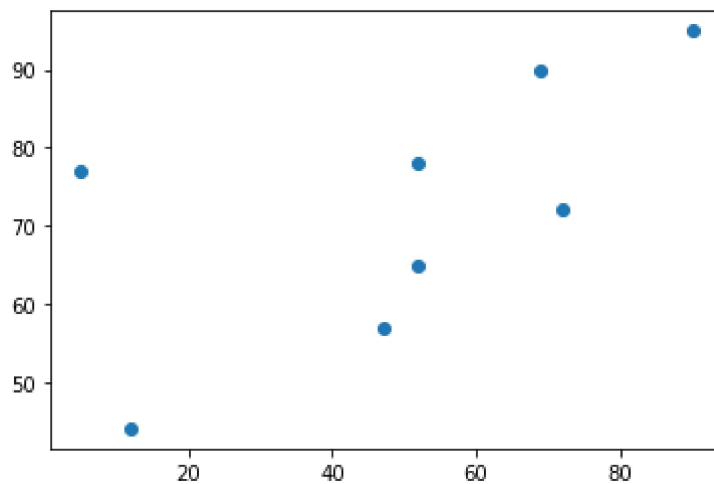
```
In [22]: plt.hist(df2['writing score'])
          plt.xlabel('writing score')
          plt.ylabel('frequency')
          plt.title("Histogram for writing score")
```

Out[22]: Text(0.5, 1.0, 'Histogram for writing score')



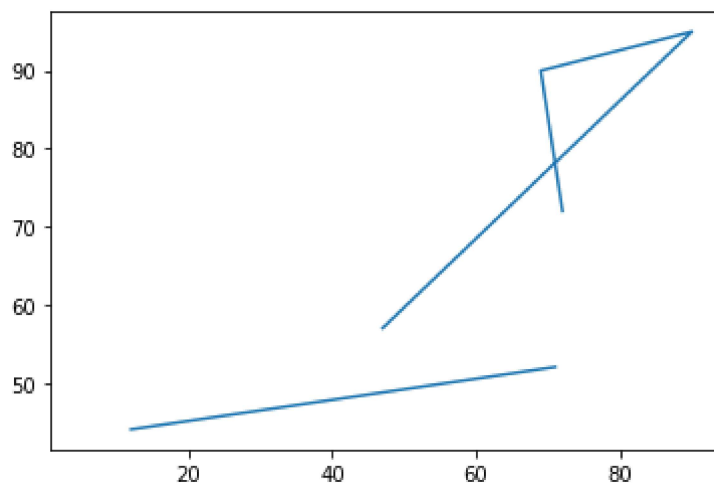
In [23]: `plt.scatter(df2["math score"], df["reading score"])`

Out[23]: <matplotlib.collections.PathCollection at 0x2071d9dc7c0>



In [24]: `plt.plot(df["math score"], df2["reading score"])`

Out[24]: [`matplotlib.lines.Line2D` at 0x2071da45c70>]



In []: