```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv("StudentPerformance.csv")
```

```
df
```

	Gender	Math Score	reading score	writing score	Placement Score	Placement Count	Region
0	Female	75	65	55	70	2	pune
1	Male	60	65	90	80	2	pimpri
2	Female	78	78	90	81	1	Mumbai
3	Male	67	90	95	82	2	pune
4	Female	89	98	72	83	3	pune
5	Male	90	69	82	84	3	pune
6	Male	98	73	99	85	3	pune
7	Female	79	98	99	86	1	pune
8	Male	80	75	76	87	2	pune
9	Female	81	87	80	88	3	Banglore

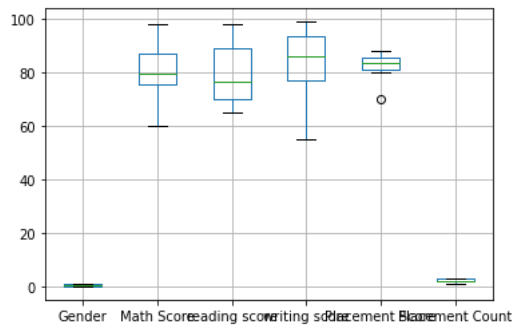
```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
df['Gender']= label_encoder.fit_transform(df['Gender'])
df['Gender'].unique()
```

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

```
df
```

	Gender	Math Score	reading score	writing score	Placement Score	Placement Count	Region
0	0	75	65	55	70	2	pune
1	1	60	65	90	80	2	pimpri
2	0	78	78	90	81	1	Mumbai
3	1	67	90	95	82	2	pune
4	0	89	98	72	83	3	pune
5	1	90	69	82	84	3	pune
6	1	98	73	99	85	3	pune
7	0	79	98	99	86	1	pune
8	1	80	75	76	87	2	pune
9	0	81	87	80	88	3	Banglore

```
import matplotlib.pyplot as plt
boxplot = df.boxplot()
plt.show()
```




```
import scipy.stats as stats
df['Math Score'].mean()
```

```
mean = df['Math Score'].mean()


df['Math Score'].std()
std = df['Math Score'].std()


zscores = stats.zscore(df['Math Score'])
print(zscores)
```

```
0   -0.447295
1   -1.874831
2   -0.161787
3   -1.208647
4    0.885072
5    0.980241
6    1.741594
7   -0.066618
8    0.028551
9    0.123720
Name: Math Score, dtype: float64
```

```
threshold = 0
outlier = []
for i in df['Math Score']:
    z = (i-mean)/std
    if z > threshold:
        outlier.append(i)
print('outlier in dataset is', outlier)
```

```
outlier in dataset is [89, 90, 98, 80, 81]
```