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

df = pd.read_csv(r"C:\Users\Jayditya\Downloads\
StudentsPerformance0.csv")
df.head()
```

```
   gender race/ethnicity parental level of education         lunch  \
0  female        group B           bachelor's degree      standard
1  female        group C                some college      standard
2  female        group B            master's degree      standard
3    male        group A          associate's degree  free/reduced
4    male        group C                some college      standard

  test preparation course  math score  reading score  writing score
0                    none          72             72             74
1               completed          69             90             88
2                    none          90             95             93
3                    none          47             57             44
4                    none          76             78             75
```

```
df.isnull().sum()
```

```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        4
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

```
# Fill missing values in 'test preparation course' column with the
mode
df['test preparation course'] = df['test preparation
course'].fillna(df['test preparation course'].mode()[0])

# Apply log transformation on 'math score'
df['log_math_score'] = np.log(df['math score']+1)
df['log_math_score']
```

```
0      4.290459
1      4.248495
2      4.510860
3      3.871201
4      4.343805
        ...
995    4.488636
```

```
996    4.143135
997    4.094345
998    4.234107
999    4.356709
Name: log_math_score, Length: 1000, dtype: float64
```

```
# Check skewness of original math score
df['math score'].skew()
```

```
-0.27893514909431694
```

```
# Check skewness of log-transformed math score
df['log_math_score'].skew()
```

```
-4.101270402189049
```

```
# Apply square root transformation on 'math score'
df['sqrt_math_score'] = np.sqrt(df['math score'])
df['sqrt_math_score'].skew()
```

```
-1.1222289710854843
```