```
import pandas as pd
import numpy as np

data = pd.read_csv('StudentsPerformance.csv')

data.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
```

```
# Remoce some unused columns
del data['gender']
del data['race/ethnicity']
del data['parental level of education']
del data['lunch']
del data['test preparation course']

data.head()
```

```
   math score  reading score  writing score
0          72             72             74
1          69             90             88
2          90             95             93
3          47             57             44
4          76             78             75
```

```
data.shape
```

```
(1000, 3)
```

```
import matplotlib.pyplot as plt
import seaborn as sns

data.describe()
```

```
       math score  reading score  writing score
count  1000.00000    1000.000000    1000.000000
mean     66.08900      69.169000      68.054000
std      15.16308      14.600192      15.195657
min       0.00000      17.000000      10.000000
25%      57.00000      59.000000      57.750000
```

```
50%      66.00000      70.000000      69.000000
75%      77.00000      79.000000      79.000000
max     100.00000     100.000000     100.000000
```

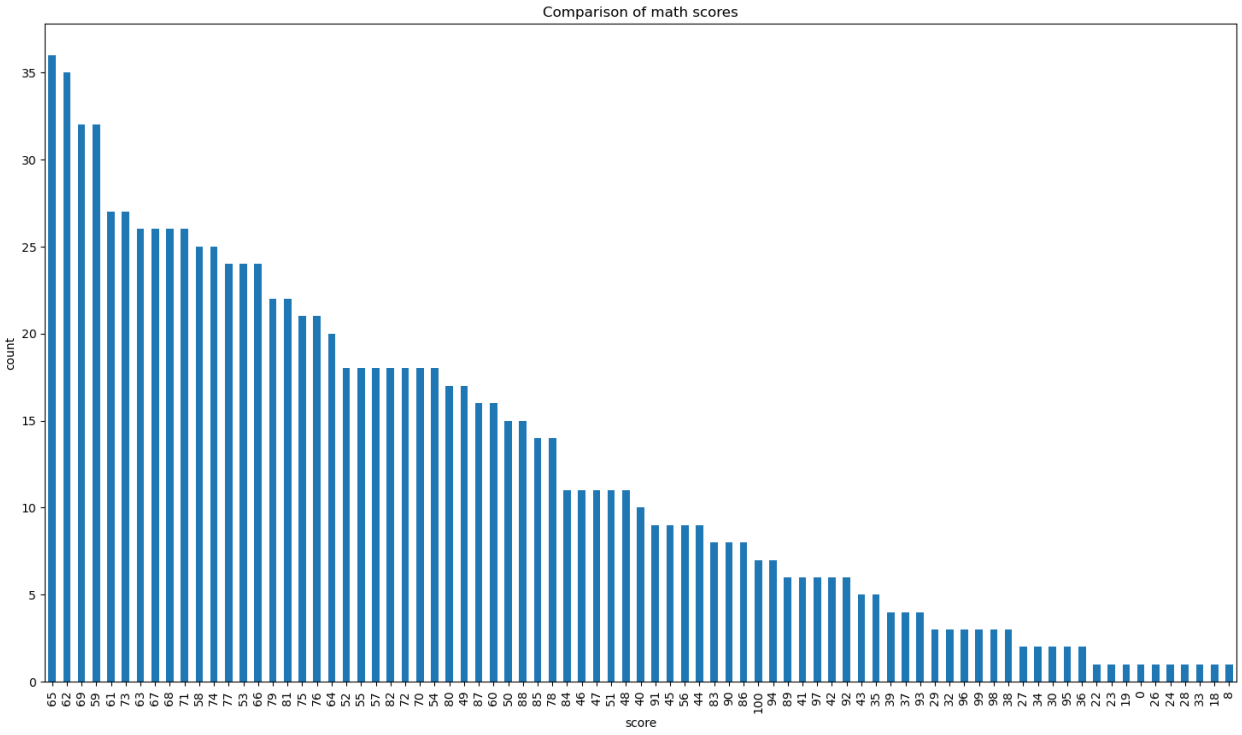
```
data.sum()
```

```
math score       66089
reading score    69169
writing score    68054
dtype: int64
```

```
data.mean()
```

```
math score       66.089
reading score    69.169
writing score    68.054
dtype: float64
```

```
# visualizing maths score
data['math score'].value_counts(normalize = True)
data['math score'].value_counts(dropna = False).plot.bar(figsize =
(18, 10))
plt.title('Comparison of math scores')
plt.xlabel('score')
plt.ylabel('count')
plt.show()
```



```
# visualizing reading score score
data['reading score'].value_counts(normalize = True)
data['reading score'].value_counts(dropna = False).plot.bar(figsize =
(18, 10), color = 'orange')
plt.title('Comparison of reading scores')
plt.xlabel('score')
plt.ylabel('count')
plt.show()
```