

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
from sklearn.preprocessing import MinMaxScaler
```

```
[2]: df = pd.read_csv("StudentsPerformance.csv")
```

```
[3]: df.head()
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

```
[4]: df.isnull()
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
995	False	False	False	False	False	False	False	False
996	False	False	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False
998	False	False	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False

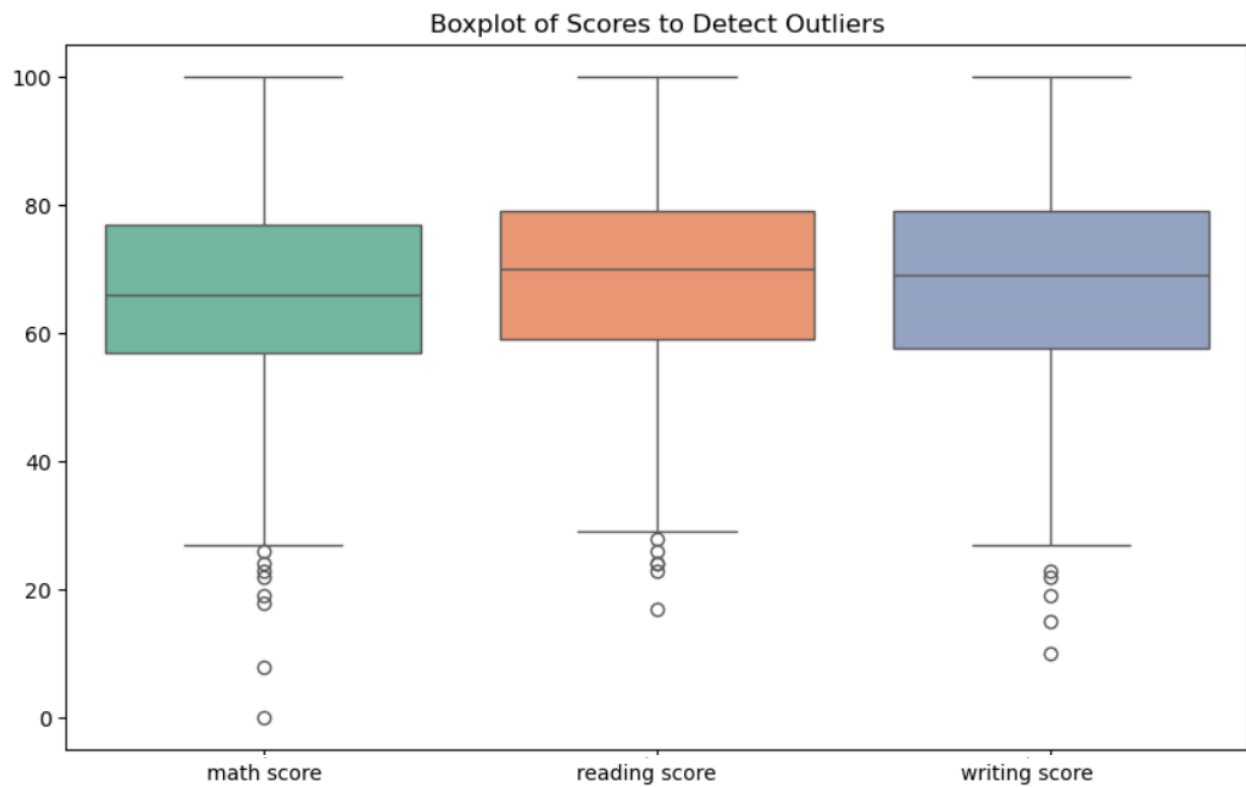
1000 rows × 8 columns

```
[5]: df.isnull().sum()
```

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

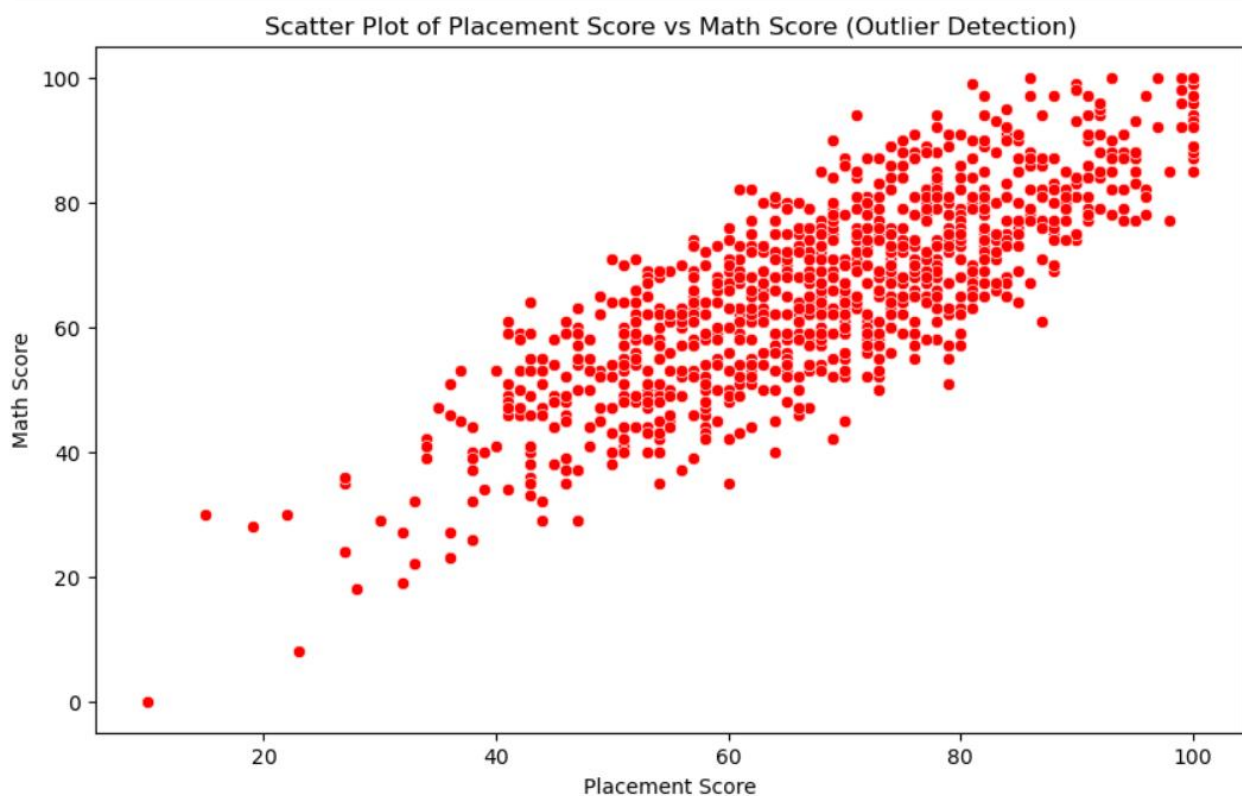
detect outliers

```
[8]: plt.figure(figsize=(10, 6))
sns.boxplot(data=df[['math score', 'reading score', 'writing score']], palette='Set2')
plt.title("Boxplot of Scores to Detect Outliers")
plt.show()
```



with scatterplot

```
•[17]: plt.figure(figsize=(10, 6))
sns.scatterplot(x=df['writing score'], y=df['math score'], color='red')
plt.title("Scatter Plot of Placement Score vs Math Score (Outlier Detection)")
plt.xlabel("Writing score")
plt.ylabel("Math Score")
plt.show()
```



with iqr

```
[20]: Q1 = df['reading score'].quantile(0.25)
      Q3 = df['reading score'].quantile(0.75)
      IQR = Q3 - Q1
      lower_bound = Q1 - 1.5 * IQR
      upper_bound = Q3 + 1.5 * IQR
```

```
[22]: lower_bound
```

```
[22]: 29.0
```

```
[24]: upper_bound
```

```
[24]: 109.0
```

histogram

```
[30]: plt.figure(figsize=(12, 5))

      plt.subplot(1, 2, 1)
      plt.hist(df_no_outliers['math score'], bins=10, color='orange', alpha=0.7)
      plt.title("Before Log Transformation")

[30]: Text(0.5, 1.0, 'Before Log Transformation')
```

