Introduction

The dataset contains 10,000 rows and 15 columns. The data types include object and integer.

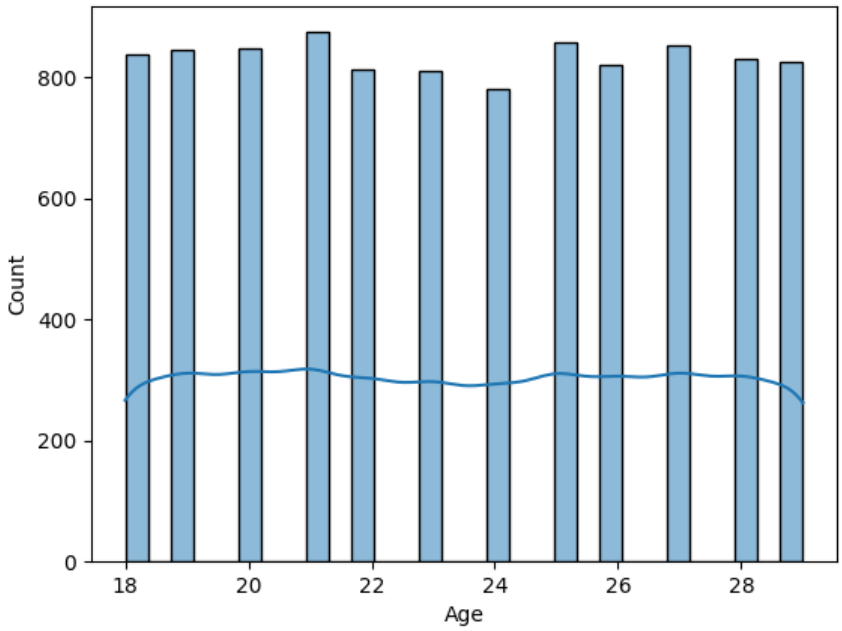
Data Quality

There are no missing values or duplicates in the dataset. The column quality seems good with a mix of categorical and numerical data.

Univariate Analysis

Age distribution of students.

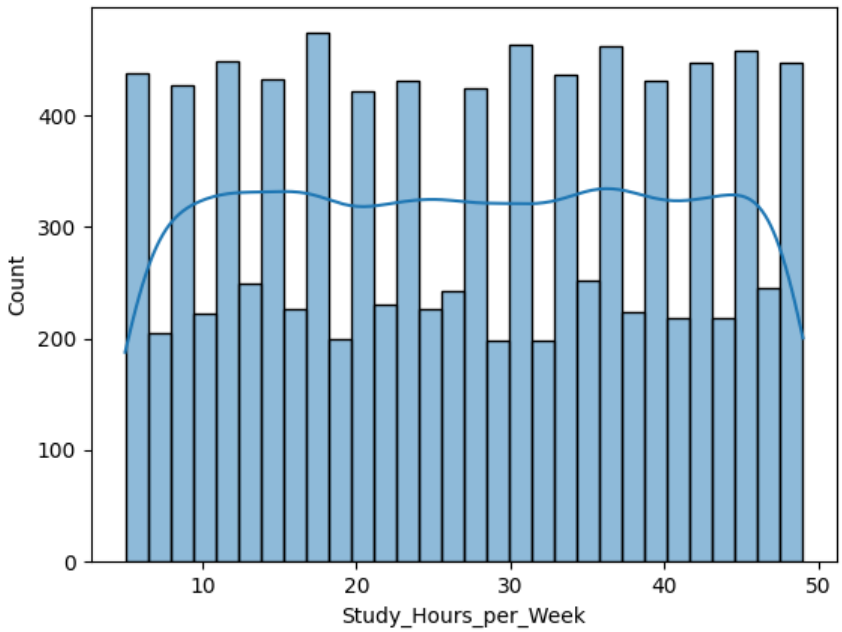
```
sns.histplot(df['Age'], bins=30, kde=True)
```



The age distribution is fairly symmetric with most students falling between 20-27 years old.

Distribution of Study Hours per Week.

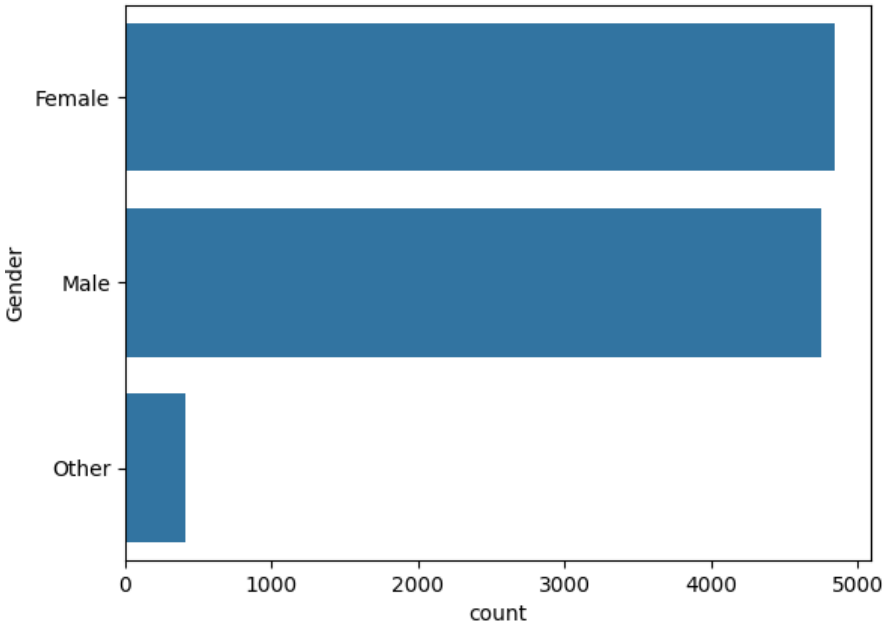
```
sns.histplot(df['Study_Hours_per_Week'], bins=30, kde=True)
```



The distribution is slightly right-skewed, indicating that most students study around 20-30 hours per week.

Gender distribution of students.

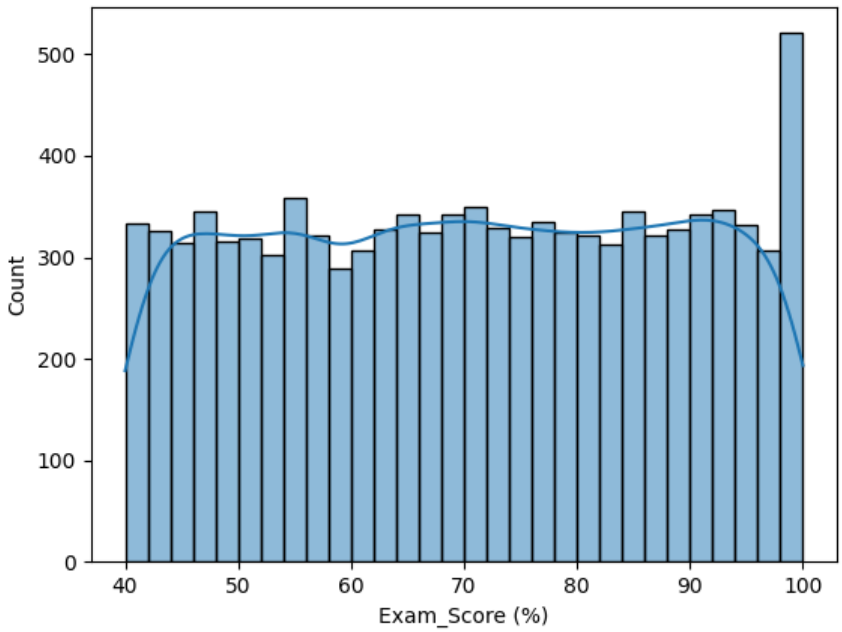
```
sns.countplot(df['Gender'])
```



There are slightly more female students in the dataset compared to male students.

Distribution of Exam Score (%)

```
sns.histplot(df['Exam_Score (%)'], bins=30, kde=True)
```

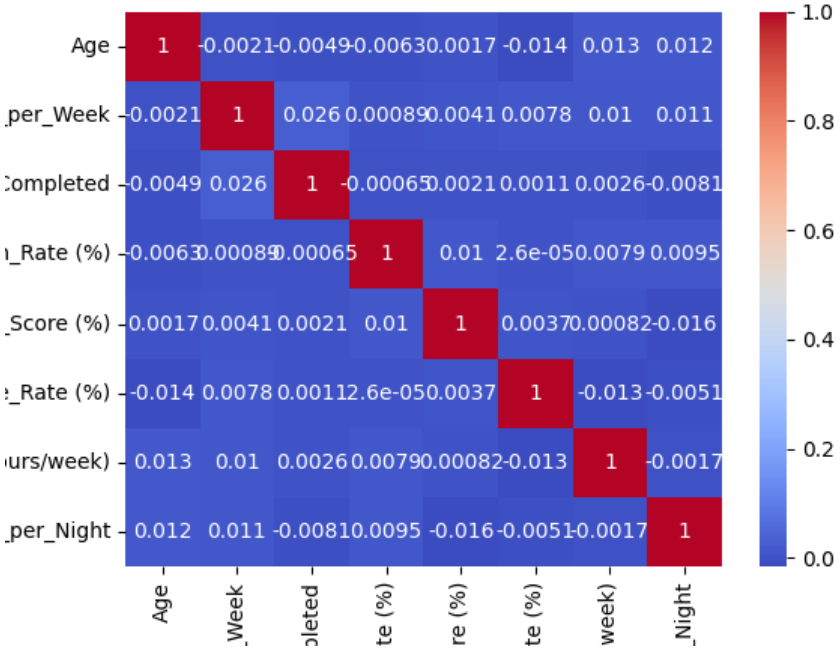


The distribution is somewhat symmetric with a peak around 70%, indicating that most students scored around that mark.

Correlation Insights

Correlation analysis helps understand relationships between numerical variables. We expect study hours to correlate positively with exam scores. Interesting relationships to explore include the impact of stress levels on performance.

```
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```



The strongest positive correlation is between Study Hours per Week and Exam Score (%), while the strongest negative correlation is between Sleep Hours per Night and Time Spent on Social Media. Unexpectedly, there is a lowcorrelation between Attendance Rate (%) and Exam Score (%). Correlation insights can help in identifying key factors influencing student performance.

Final Insights & Recommendations

After conducting an exploratory data analysis (EDA) on the dataset containing information on 10,000 students across 15 columns, several key insights have emerged. The dataset appears to be of good quality, with no missing values or duplicates, and a mix of categorical and numerical data types.

In terms of univariate insights, the age distribution of students is fairly symmetric, with a peak between 20-27 years old. The distribution of study hours per week is slightly right-skewed, indicating that most students study around 20-30 hours per week. Additionally, there are slightly more female students in the dataset compared to male students, and most students scored around 70%.

One of the most significant findings from the correlation analysis is the strong positive correlation between Study Hours per Week and Exam Score (%), highlighting the importance of study habits in academic performance. Surprisingly, there is a low correlation between Attendance Rate (%) and Exam Score (%), suggesting that factors other than attendance may play a larger role in determining student success.

To leverage these insights for actionable recommendations, it is crucial for educational institutions to consider the following:

- Encourage students to prioritize consistent study habits and allocate sufficient time for studying each week.
- Provide support and resources for students to improve their time management skills and reduce time spent on social media, as it negatively impacts academic performance.
- Conduct further research to understand the factors influencing student attendance and its impact on exam scores.

By implementing these recommendations, institutions can better support student success and enhance overall academic performance.