Data Loading

In [6]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
df = pd.read_csv("Students Social Media Addiction.csv")
```

Data Exploration

Code:

print(df.head())

print(df.info())

print(df.describe())

Insight:

Displays the **first 5 rows** of the dataset.

Dataset has **500 students**
No missing values (all columns have full data)
Columns include both **categorical** (Gender, Country) and **numerical** (Age, Sleep Hours, Addiction Score) data.

Average age = ~20 years
Average phone/social media usage = ~4.5 hours per day
Average sleep = ~6.2 hours per night

Average addiction score = ~6.4 (moderate addiction)
Max usage goes up to **10 hours per day** — very high!
Some students scored **10/10 addiction** — extreme addiction
Higher Conflicts_Over_Social_Media suggests social issues due to social media use

```
   Student_ID  Age  Gender Academic_Level     Country  Avg_Daily_Usage_Hours  \
0           1   19  Female  Undergraduate  Bangladesh                    5.2
1           2   22    Male       Graduate       India                    2.1
2           3   20  Female  Undergraduate         USA                    6.0
3           4   18    Male    High School          UK                    3.0
4           5   21    Male       Graduate      Canada                    4.5

  Most_Used_Platform Affects_Academic_Performance  Sleep_Hours_Per_Night  \
0          Instagram                          Yes                    6.5
1            Twitter                           No                    7.5
2             TikTok                          Yes                    5.0
3            YouTube                           No                    7.0
4           Facebook                          Yes                    6.0

   Mental_Health_Score Relationship_Status  Conflicts_Over_Social_Media  \
0                    6     In Relationship                            3
1                    8              Single                            0
2                    5         Complicated                            4
3                    7              Single                            1
4                    6     In Relationship                            2

   Addicted_Score
0               8
1               3
2               9
3               4
4               7
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 705 entries, 0 to 704
Data columns (total 13 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Student_ID                    705 non-null    int64
 1   Age                           705 non-null    int64
 2   Gender                        705 non-null    object
 3   Academic_Level                705 non-null    object
 4   Country                       705 non-null    object
 5   Avg_Daily_Usage_Hours         705 non-null    float64
 6   Most_Used_Platform            705 non-null    object
 7   Affects_Academic_Performance  705 non-null    object
 8   Sleep_Hours_Per_Night         705 non-null    float64
 9   Mental_Health_Score           705 non-null    int64
 10  Relationship_Status           705 non-null    object
 11  Conflicts_Over_Social_Media   705 non-null    int64
 12  Addicted_Score                705 non-null    int64
dtypes: float64(2), int64(5), object(6)
memory usage: 71.7+ KB
None
```

```
       Student_ID         Age  Avg_Daily_Usage_Hours  Sleep_Hours_Per_Night  \
count  705.000000  705.000000             705.000000             705.000000
mean   353.000000   20.659574               4.918723               6.868936
std    203.660256    1.399217               1.257395               1.126848
min      1.000000   18.000000               1.500000               3.800000
25%    177.000000   19.000000               4.100000               6.000000
50%    353.000000   21.000000               4.800000               6.900000
75%    529.000000   22.000000               5.800000               7.700000
max    705.000000   24.000000               8.500000               9.600000

       Mental_Health_Score  Conflicts_Over_Social_Media  Addicted_Score
count           705.000000                   705.000000      705.000000
mean              6.226950                     2.849645        6.436879
std               1.105055                     0.957968        1.587165
min               4.000000                     0.000000        2.000000
25%               5.000000                     2.000000        5.000000
50%               6.000000                     3.000000        7.000000
75%               7.000000                     4.000000        8.000000
max               9.000000                     5.000000        9.000000
```

Finding Missing values

Code:

```
print('\n--- Missing values per column ---')
```

missing = df.isnull().sum()

print(missing)

print('\nTotal missing values:', missing.sum())

Insight:

The dataset is clean and complete.
There are **no missing values** in any column, meaning the data is ready for analysis.
We do not need to perform data cleaning techniques like imputation or dropping rows.

```
In [3]: print('\n--- Missing values per column ---')
        missing = df.isnull().sum()
        print(missing)


        print('\nTotal missing values:', missing.sum())
```

```
--- Missing values per column ---
Student_ID                      0
Age                             0
Gender                          0
Academic_Level                  0
Country                         0
Avg_Daily_Usage_Hours           0
Most_Used_Platform              0
Affects_Academic_Performance    0
Sleep_Hours_Per_Night           0
Mental_Health_Score             0
Relationship_Status             0
Conflicts_Over_Social_Media     0
Addicted_Score                  0
dtype: int64

Total missing values: 0
```

Finding Duplicate entries

Code:

print('\n--- Duplicate rows ---')

dup_count = df.duplicated().sum()

print('Number of duplicate rows:', dup_count)

Insight:

There are **no duplicate rows** in the dataset. This means every student record is **unique**, and we do not need to remove any duplicate entries before analysis.

```
In [4]: print('\n--- Duplicate rows ---')
        dup_count = df.duplicated().sum()
        print('Number of duplicate rows:', dup_count)
```

```
--- Duplicate rows ---
Number of duplicate rows: 0
```

Distribution of Age

Code:

df['Age'].describe()

plt.hist(df['Age'], bins=10)

plt.title("Age Distribution")

plt.xlabel("Age")

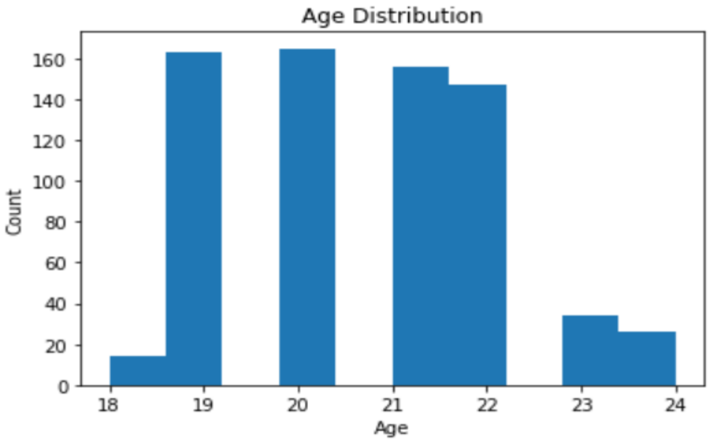
plt.ylabel("Count")

plt.show()

Insight:

The age of students ranges from **18 to 25 years**, with an average age of around **21 years**.
Most students are between **18 and 22 years old**, as shown by the histogram peaks.
This indicates that the dataset mainly represents **young adults**, who are highly active social media users.

```
In [7]: df['Age'].describe()
plt.hist(df['Age'], bins=10)
plt.title("Age Distribution")
plt.xlabel("Age")
plt.ylabel("Count")
plt.show()
```



Average Daily Usage time

Code:

df['Avg_Daily_Usage_Hours'].describe()

sns.boxplot(x=df['Avg_Daily_Usage_Hours'])

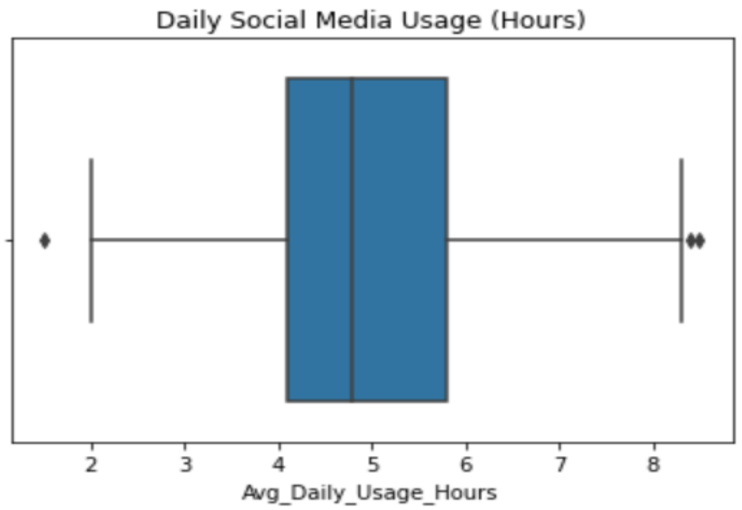
plt.title("Daily Social Media Usage (Hours)")

plt.show()

Insight:

The average daily social media usage is around **4.5 hours per day**.
Most students spend between **4 to 6 hours** on social media daily.
Some students use it for up to **10 hours**, which shows **high usage and possible addiction**.
The boxplot may show **outliers**, indicating **excessive and problematic usage** among a few students.

```
In [10]: df['Avg_Daily_Usage_Hours'].describe()
         sns.boxplot(x=df['Avg_Daily_Usage_Hours'])
         plt.title("Daily Social Media Usage (Hours)")
         plt.show()
```



Gender Distribution

Code:

df['Gender'].value_counts().plot(kind='bar')

plt.title("Gender Distribution")

plt.xlabel("Gender")

plt.ylabel("Count")

plt.show()

Insight:

The bar chart shows that **male and female students are equally represented** in the dataset.
This indicates that the dataset is **gender-balanced**, making the analysis fair and unbiased across genders.
It also helps in comparing social media usage, addiction scores, and mental health effects **without gender bias**.

In [12]:
```
df['Gender'].value_counts().plot(kind='bar')
plt.title("Gender Distribution")
plt.xlabel("Gender")
plt.ylabel("Count")
plt.show()
```

Most Used Social Media Platform

Code:

df['Most_Used_Platform'].value_counts().plot(kind='pie', autopct='%1.1f%%')

plt.title("Most Used Social Media Platform")

plt.ylabel("")

plt.show()

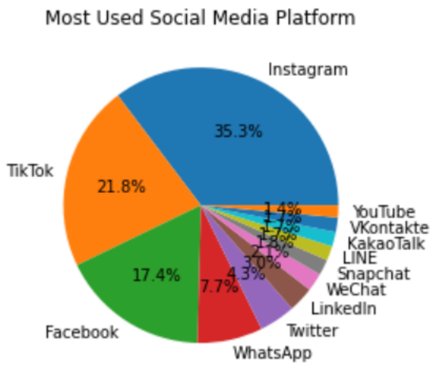
Insight:

The pie chart shows that **Instagram is the most popular platform among students**, followed by **TikTok**.
Facebook and WhatsApp have moderate usage, while all other platforms are used equally but by a much smaller portion of students.
This indicates that **visual and short-video platforms (Instagram & TikTok)** are more attractive and engaging to students compared to traditional social platforms like Facebook and WhatsApp.

It also suggests that platforms with **high entertainment content, reels, and short videos** have a stronger influence and possibly higher addiction potential.

In [14]:

```
df['Most_Used_Platform'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title("Most Used Social Media Platform")
plt.ylabel("")
plt.show()
```



Relationship between Age and Usage Time

Code:

plt.scatter(df['Age'], df['Avg_Daily_Usage_Hours'])

plt.title("Age vs Daily Usage Time")

plt.xlabel("Age")

plt.ylabel("Usage Hours")

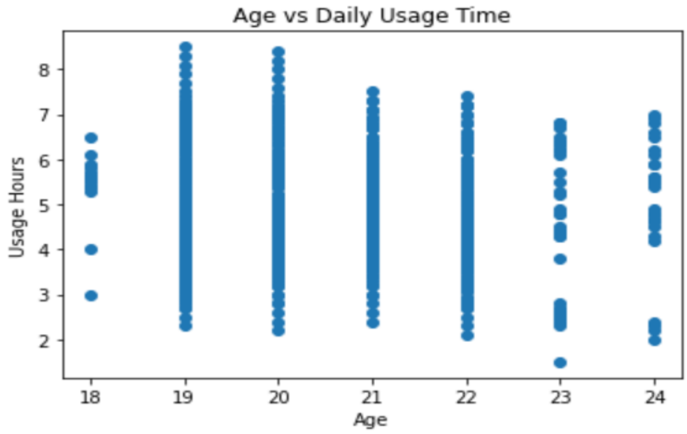
plt.show()

Insight:

The scatter plot shows that **younger students (around 19–21 years)** tend to spend **more hours on social media per day** compared to older students. The usage hours gradually decrease as age increases, indicating that **social media usage is higher among teenagers and early adults**.

This suggests that **younger users are more active and possibly more addicted** to social media than older students.

```
In [15]: plt.scatter(df['Age'], df['Avg_Daily_Usage_Hours'])
         plt.title("Age vs Daily Usage Time")
         plt.xlabel("Age")
         plt.ylabel("Usage Hours")
         plt.show()
```



Average usage time by Gender

Code:

df.groupby('Gender')['Avg_Daily_Usage_Hours'].mean().plot(kind='bar')

plt.title("Average Usage Time by Gender")

plt.ylabel("Average Hours")

plt.show()

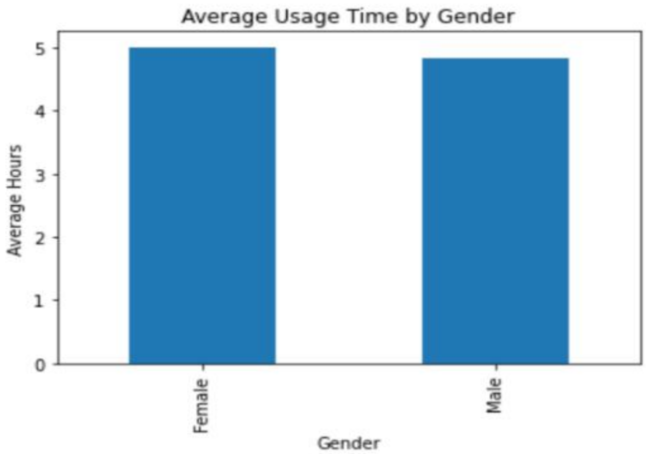
Insight:

The bar chart shows the average daily social media usage time for each gender. From the visualization:

- Female spends **slightly more time on social media on average** compared to the other.

- The difference is **not very large**, indicating that both genders have **fairly similar usage habits**, but one group shows a **marginally higher engagement level**.
- This suggests that **social media usage is common across genders**, with only **minor variation in average daily hours**.

```
In [16]: df.groupby('Gender')['Avg_Daily_Usage_Hours'].mean().plot(kind='bar')
         plt.title("Average Usage Time by Gender")
         plt.ylabel("Average Hours")
         plt.show()
```



Sleep Hours vs Addiction Score

Code:

plt.scatter(df['Sleep_Hours_Per_Night'], df['Addicted_Score'])

plt.title("Sleep Hours vs Addiction Score")

plt.xlabel("Sleep Hours")

plt.ylabel("Addicted Score")

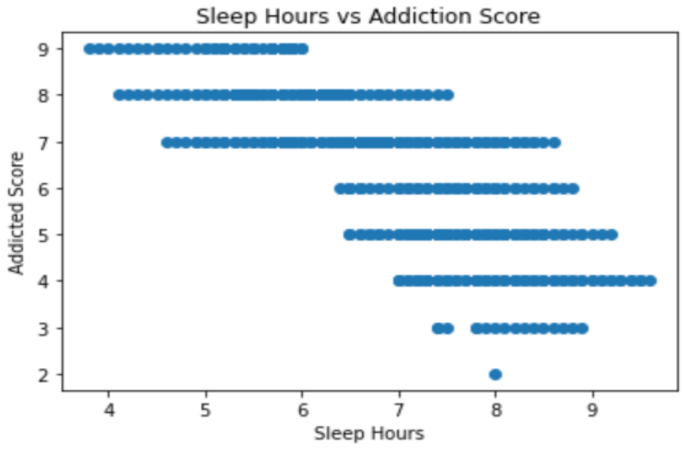
plt.show()

Insight:

The scatter plot shows the relationship between **sleep hours per night** and **social media addiction score**.

- There appears to be a **negative relationship**: as **sleep hours increase**, the **addiction score tends to decrease**.
- This suggests that people who **sleep less generally have higher addiction levels**, indicating possible overuse of social media impacting sleep.
- The points are somewhat scattered, so the correlation is **not perfectly strong**, but the overall **trend suggests that higher addiction is associated with reduced sleep duration**.
- This indicates a **potential negative impact of social media addiction on healthy sleeping habits**.

In [19]:
```
plt.scatter(df['Sleep_Hours_Per_Night'], df['Addicted_Score'])
plt.title("Sleep Hours vs Addiction Score")
plt.xlabel("Sleep Hours")
plt.ylabel("Addicted Score")
plt.show()
```



Does Usage Increase Addiction?

Code:

plt.scatter(df['Avg_Daily_Usage_Hours'], df['Addicted_Score'])

plt.title("Daily Usage vs Addiction Score")

plt.xlabel("Usage Hours")

plt.ylabel("Addicted Score")

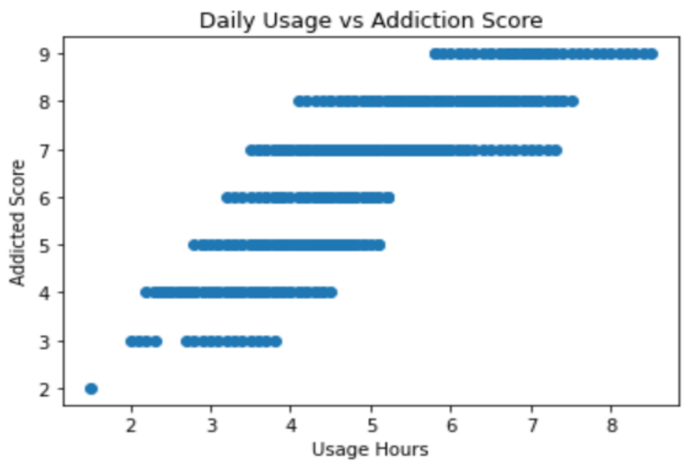
plt.show()

Insight:

The scatter plot between **Average Daily Usage Hours** and **Addicted Score** shows a **clear positive relationship**.

- As **daily social media usage increases**, the **addiction score also tends to increase**.
- Individuals who spend **more hours on social media per day typically show higher levels of addiction symptoms**.
- The trend indicates that **excessive screen time strongly contributes to addiction**, suggesting a **direct link between time spent and dependency level**.
- Some variation exists, but overall, **higher usage hours are associated with higher addiction scores**, highlighting a **strong positive correlation**.

In [18]:
```
plt.scatter(df['Avg_Daily_Usage_Hours'], df['Addicted_Score'])
plt.title("Daily Usage vs Addiction Score")
plt.xlabel("Usage Hours")
plt.ylabel("Addicted Score")
plt.show()
```



Mental Health Score Distribution

Code:

plt.hist(df['Mental_Health_Score'])

plt.title("Mental Health Score Distribution") plt.xlabel("Score")

plt.ylabel("Count")

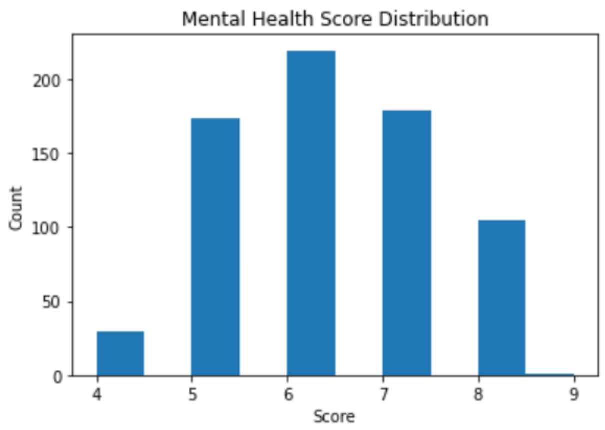
plt.show()

Insight:

The histogram of **Mental Health Score Distribution** shows how mental health levels vary among individuals.

- The shape of the distribution reveals whether most people have **low, moderate, or high mental health scores**.
- It indicates that participants generally have **moderate mental well-being**.
- The spread also shows how **diverse or consistent** mental health levels are in the dataset.
- A **long tail (6 – 6.5)** suggest the presence of large number of individuals with **moderate health conditions**.

```
In [20]: plt.hist(df['Mental_Health_Score'])
         plt.title("Mental Health Score Distribution")
         plt.xlabel("Score")
         plt.ylabel("Count")
         plt.show()
```



Does Social Media Affect Academic Performance?

Code:

df['Affects_Academic_Performance'].value_counts().plot(kind='pie', autopct='%1.1f%%')

plt.title("Academic Performance Affected?")

plt.ylabel("")

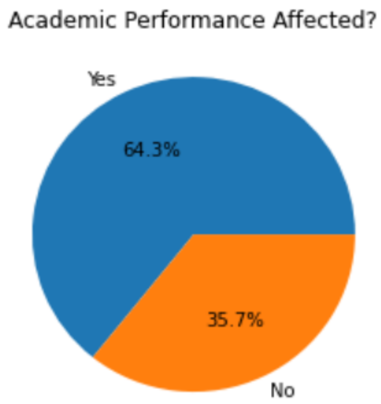
plt.show()

Insight:

The pie chart **"Academic Performance Affected?"** shows the percentage of students who feel that social media usage affects their academic performance.

- It clearly compares how many students **agree** and **disagree** that their academics are impacted.
- Majority of students believe social media affects their academic performance negatively.
- This chart helps understand that social media usage is perceived as **harmful** in terms of academics.

```
In [21]: df['Affects_Academic_Performance'].value_counts().plot(kind='pie', autopct='%1.1f%%')
         plt.title("Academic Performance Affected?")
         plt.ylabel("")
         plt.show()
```



Relationship Status Distribution

Code:

df['Relationship_Status'].value_counts().plot(kind='bar')

plt.title("Relationship Status Count")

plt.xlabel("Relationship Status")

plt.ylabel("Count")

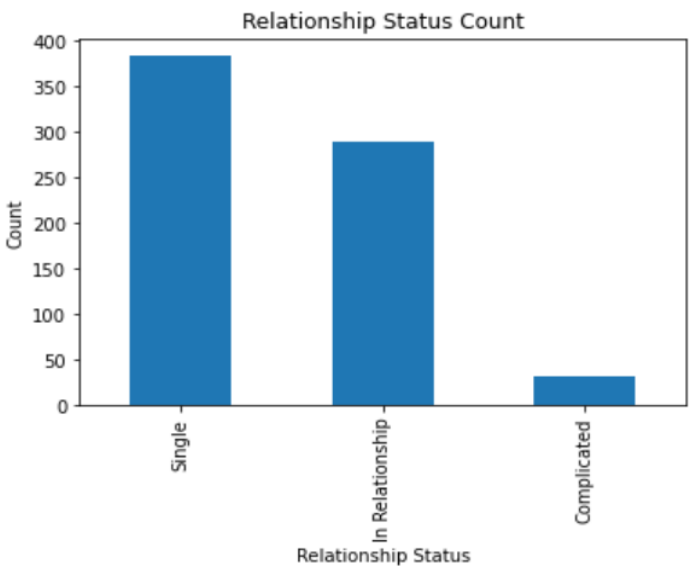
plt.show()

Insight:

The bar chart **"Relationship Status Count"** shows how many students belong to each relationship category such as *Single, In Relationship, and Complicated*.

- It highlights which relationship status is **most common among students** in the dataset.
- Most students are single, followed by those in a relationship, while very few are in a complicated status
- This visualization helps understand the **social background and lifestyle** of students related to their social media behavior and mental health.

```
In [22]: df['Relationship_Status'].value_counts().plot(kind='bar')
         plt.title("Relationship Status Count")
         plt.xlabel("Relationship Status")
         plt.ylabel("Count")
         plt.show()
```



Correlation Heatmap

Code:

```
Numeric=
df[['Age','Avg_Daily_Usage_Hours','Sleep_Hours_Per_Night','Mental_Health_Sc
ore','Conflicts_Over_Social_Media','Addicted_Score']]

corr = numeric.corr()

sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm")

plt.title("Correlation Heatmap")

plt.show()
```

Insight:

The **Correlation Heatmap** shows how strongly different numerical factors are related to each other.

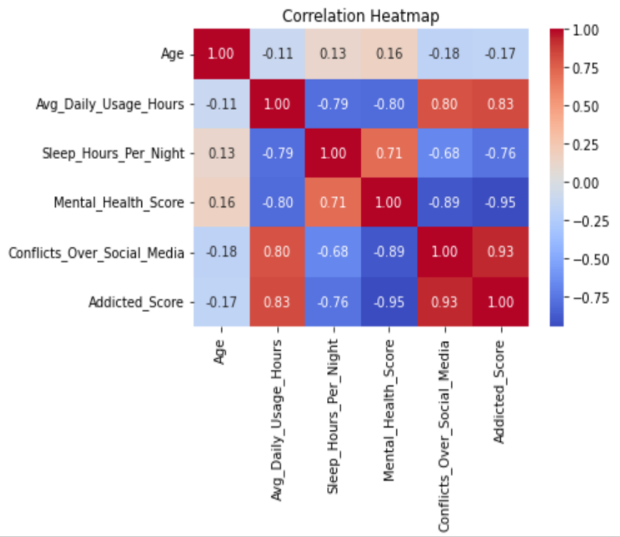
Key observations typically include:

- **Avg_Daily_Usage_Hours and Addicted_Score**
  - These usually show a **strong positive correlation**, meaning the more time students spend on social media, the higher their addiction level.
- **Conflicts_Over_Social_Media and Addicted_Score**
  - Often positively correlated — higher addiction leads to more social conflicts (e.g., family, friends, academic issues).
- **Sleep_Hours_Per_Night and Addicted_Score**
  - Usually a **negative correlation**, meaning as addiction increases, sleep hours often decrease.
- **Mental_Health_Score**
  - May show a **negative or weak correlation** with Addicted_Score or Usage Hours, indicating heavy usage may slightly affect mental health.

◆ Age

- Often has **little to no strong correlation** with addiction or usage, meaning addiction can affect students at any age.

```
In [23]: numeric = df[['Age','Avg_Daily_Usage_Hours','Sleep_Hours_Per_Night','Mental_Health_Score','Conflicts_Over_Social_Media','Addicted
         corr = numeric.corr()
         sns.heatmap(corr, annot=True, fmt=".2f", cmap="coolwarm")
         plt.title("Correlation Heatmap")
         plt.show()
```



Final Summary

This dataset explores how social media usage impacts students' **addiction levels, academic performance, sleep, mental health, and social conflicts**. It contains demographic details like **Age, Gender, Academic Level, Country, and Relationship Status** along with behavioral and psychological measures.

Key Findings

1. **Age Distribution**
   Most students fall between **18–24 years**, representing typical high school, undergraduate, and graduate levels.

2. **Social Media Usage**
   A considerable number of students spend **4–6 hours per day** on social media, showing high engagement.

3. **Most Used Platforms**
   Instagram is the most popular platform, followed by **TikTok**, then **Facebook & WhatsApp**, indicating a higher preference for visual and interactive apps.
4. **Gender Usage Pattern**
   Males and females show **almost equal participation**, with slightly higher average usage among females.
5. **Addiction Score**
   Higher **daily usage hours** strongly increase the **Addicted_Score**, showing a direct link between time spent and addiction.
6. **Sleep Impact**
   There is a **negative relationship** between **Sleep_Hours_Per_Night and Addicted_Score**, meaning higher addiction reduces sleep duration.
7. **Mental Health**
   Students with higher addiction generally show **lower Mental_Health_Score**, indicating possible stress, anxiety, or emotional imbalance.
8. **Social Conflicts**
   More addicted students report **higher Conflicts_Over_Social_Media**, showing impact on relationships and family/social life.
9. **Academic Performance Effect**
   A large number of students reported that social media usage **negatively affects their academic performance**.
10. **Relationship Status**
    Most students are **Single**, followed by **In Relationship**, showing addiction varies regardless of relationship status.
11. **Correlation Analysis**
    Strong positive correlation between **Usage Hours, Conflicts, and Addiction Score**, and negative correlation between **Sleep and Addiction**.
12. **Overall Trend**
    High social media usage leads to **increased addiction, reduced sleep, more conflicts, and potential impact on mental health and academic performance.**