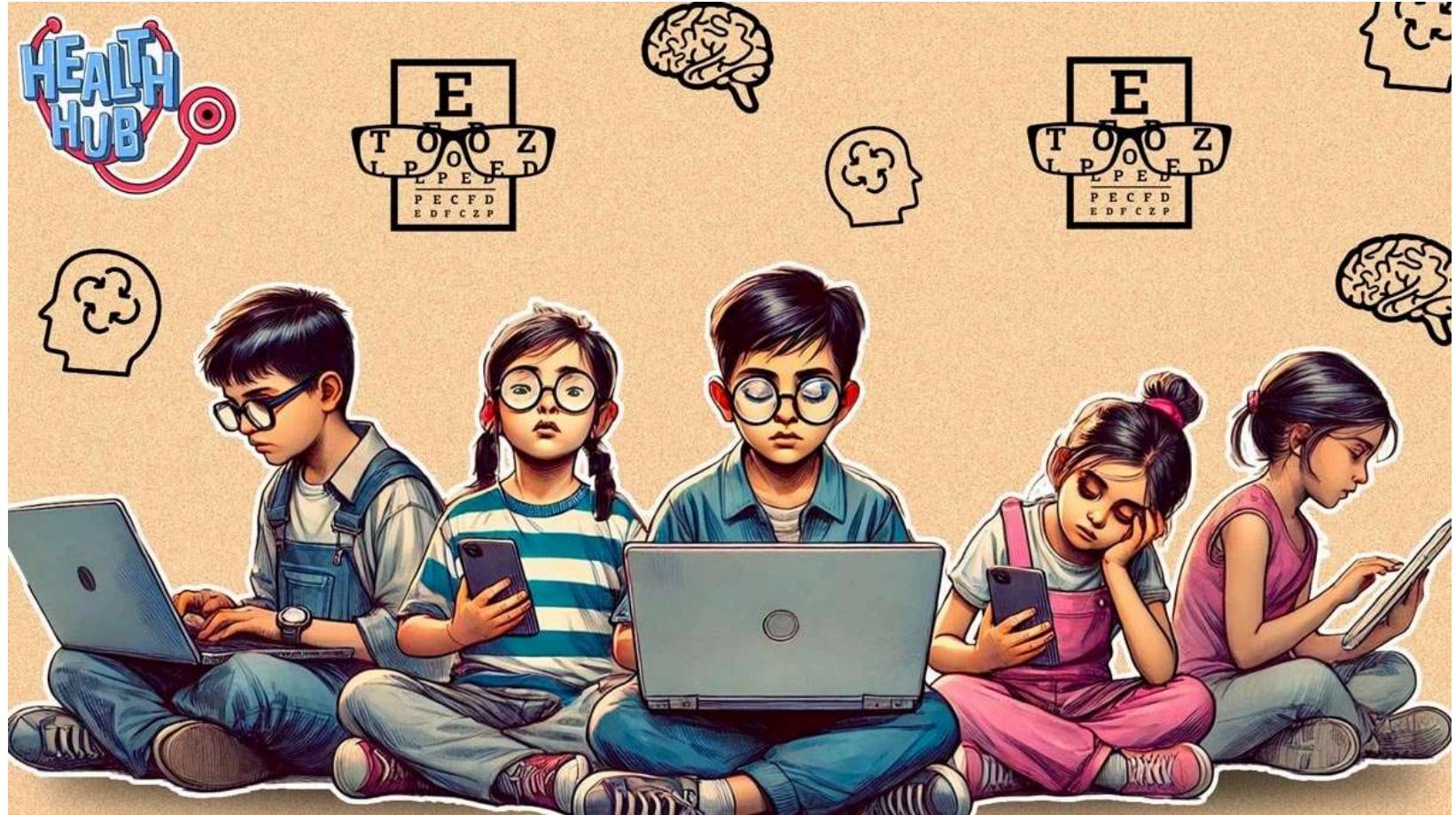


Indian Kids Screentime data Analysis



About Dataset

This dataset simulates screen time patterns of 9712 Indian children aged 8 to 18 years, built using real-world trends and scientific studies conducted in India between 2023–2024. It combines urban and rural demographics, reflecting differences in device access, screen habits, and health outcomes.

Screen time is broken down by:

Age and gender

Primary screen device (e.g., smartphone, TV)

Time split between educational and recreational use

Whether screen time exceeds Indian Academy of Pediatrics (IAP) guidelines

Likely health impacts (e.g., poor sleep, eye strain, anxiety)



Understanding Screentime Impact: An Analytical Study on Indian Kids (2025)

In [1]:

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

import warnings
#warnings.filterwarnings()
```

step-by-step data analysis operations

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

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

```
Out[3]:   Age  Gender  Avg_Daily_Screen_Time_hr  Primary_Device  Exceeded_Recommended_Limit  Educational_to_Recreational_Ratio  Health_Impacts
0    14    Male            3.99      Smartphone             True                0.42
1    11  Female            4.61        Laptop             True                0.30
2    18  Female            3.73          TV             True                0.32
3    15  Female            1.21        Laptop            False                0.39
4    12  Female            5.89      Smartphone             True                0.49
```



```
In [4]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9712 entries, 0 to 9711
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Age              9712 non-null    int64  
 1   Gender            9712 non-null    object  
 2   Avg_Daily_Screen_Time_hr  9712 non-null    float64
 3   Primary_Device    9712 non-null    object  
 4   Exceeded_Recommended_Limit  9712 non-null    bool   
 5   Educational_to_Recreational_Ratio  9712 non-null    float64
 6   Health_Impacts    6494 non-null    object  
 7   Urban_or_Rural    9712 non-null    object  
dtypes: bool(1), float64(2), int64(1), object(4)
memory usage: 540.7+ KB
```

⚠️ Warnings from the Analysis

1. **Excessive ScreenTime Harms Health** Prolonged exposure to screens is linked to eye strain, poor posture, disrupted sleep cycles, and increased risk of obesity due to reduced physical activity.
2. **Mental and Social Well-Being at Risk** Higher screen time correlates with symptoms of anxiety, reduced attention span, and lower emotional regulation, especially among pre-teens and teenagers.
3. **Academic Distraction** While screens are used for learning, unregulated use leads to distractions, reduced academic performance, and dependency on digital shortcuts.
4. **Loss of Outdoor Activities** Children spending more than 3 hours per day on screens are significantly less likely to engage in outdoor play and physical interaction, crucial for growth and motor skill development.
5. **Digital Addiction and Dependency** Patterns indicate that without parental controls or routines, many children show signs of digital dependency, such as irritability when not using devices.
6. **Cybersecurity & Exposure Risks** Increased unsupervised online activity raises the risk of children encountering inappropriate content, online predators, or developing unsafe online habits.

Data Cleaning

```
In [5]: df.isNull()
```

Out[5]:

	Age	Gender	Avg_Daily_Screen_Time_hr	Primary_Device	Exceeded_Recommended_Limit	Educational_to_Recreational_Ratio
0	False	False		False	False	False
1	False	False		False	False	False
2	False	False		False	False	False
3	False	False		False	False	False
4	False	False		False	False	False
...
9707	False	False		False	False	False
9708	False	False		False	False	False
9709	False	False		False	False	False
9710	False	False		False	False	False
9711	False	False		False	False	False

9712 rows × 8 columns



In [6]: `df.columns`

Out[6]: `Index(['Age', 'Gender', 'Avg_Daily_Screen_Time_hr', 'Primary_Device', 'Exceeded_Recommended_Limit', 'Educational_to_Recreational_Ratio', 'Health_Impacts', 'Urban_or_Rural'], dtype='object')`

In [7]: `df.describe()`

```
Out[7]:
```

	Age	Avg_Daily_Screen_Time_hr	Educational_to_Recreational_Ratio
count	9712.000000	9712.000000	9712.000000
mean	12.979201	4.352837	0.427226
std	3.162437	1.718232	0.073221
min	8.000000	0.000000	0.300000
25%	10.000000	3.410000	0.370000
50%	13.000000	4.440000	0.430000
75%	16.000000	5.380000	0.480000
max	18.000000	13.890000	0.600000

```
In [8]: df.isnull().sum()
```

```
Out[8]: Age                      0  
Gender                     0  
Avg_Daily_Screen_Time_hr    0  
Primary_Device                0  
Exceeded_Recommended_Limit    0  
Educational_to_Recreational_Ratio 0  
Health_Impacts            3218  
Urban_or_Rural                  0  
dtype: int64
```

```
In [9]: df.dtypes
```

```
Out[9]: Age                  int64  
Gender                 object  
Avg_Daily_Screen_Time_hr float64  
Primary_Device          object  
Exceeded_Recommended_Limit bool  
Educational_to_Recreational_Ratio float64  
Health_Impacts          object  
Urban_or_Rural           object  
dtype: object
```

Handle Missing Data

```
In [10]: # Calculate mode of the Health_Impacts column  
mode_value = df["Health_Impacts"].mode()[0]  
  
# Fill missing values with the mode  
df["Health_Impacts"].fillna(mode_value, inplace=True)
```

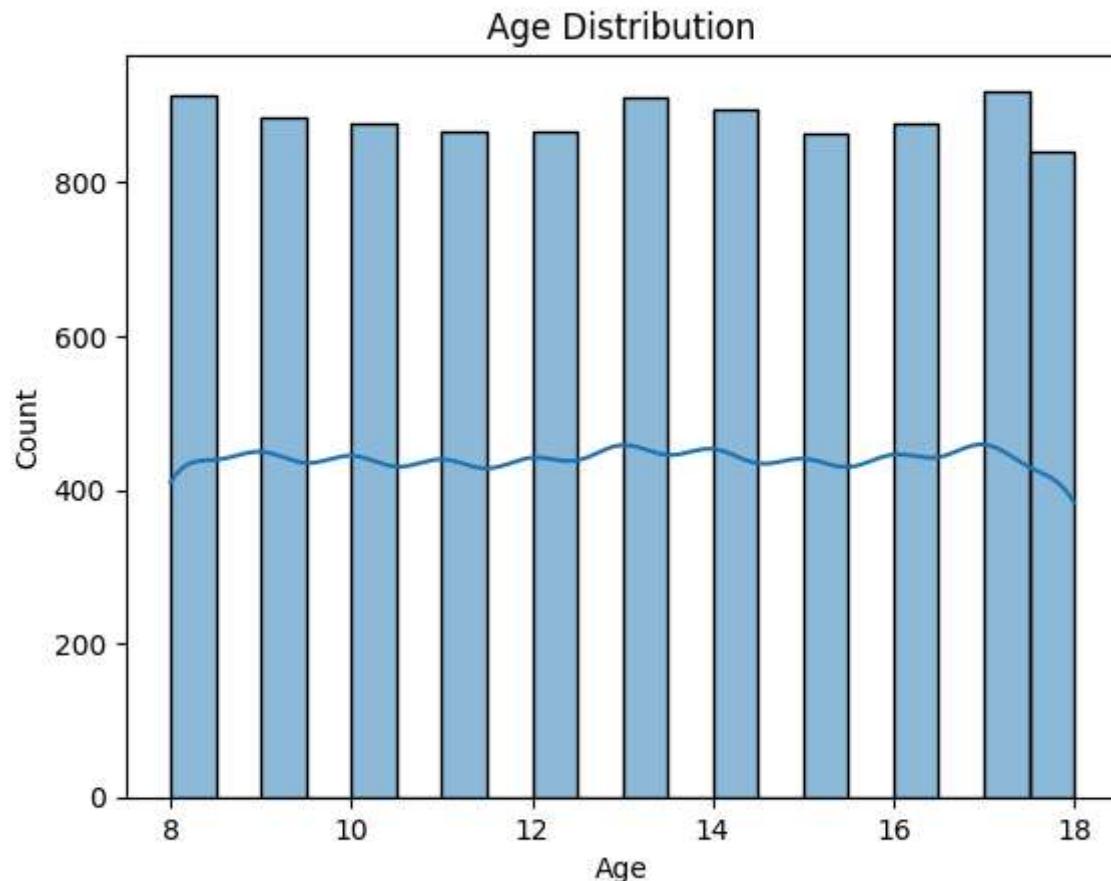
C:\Users\hi\AppData\Local\Temp\ipykernel_16264\724824926.py:5: FutureWarning: A value is trying to be set on a copy of a DataFrame or Series through chained assignment using an `inplace` method.
The behavior will change in pandas 3.0. This `inplace` method will never work because the intermediate object on which we are setting values always behaves as a copy.

For example, when doing '`df[col].method(value, inplace=True)`', try using '`df.method({col: value}, inplace=True)`' or `df[col] = df[col].method(value)` instead, to perform the operation `inplace` on the original object.

```
df["Health_Impacts"].fillna(mode_value, inplace=True)
```

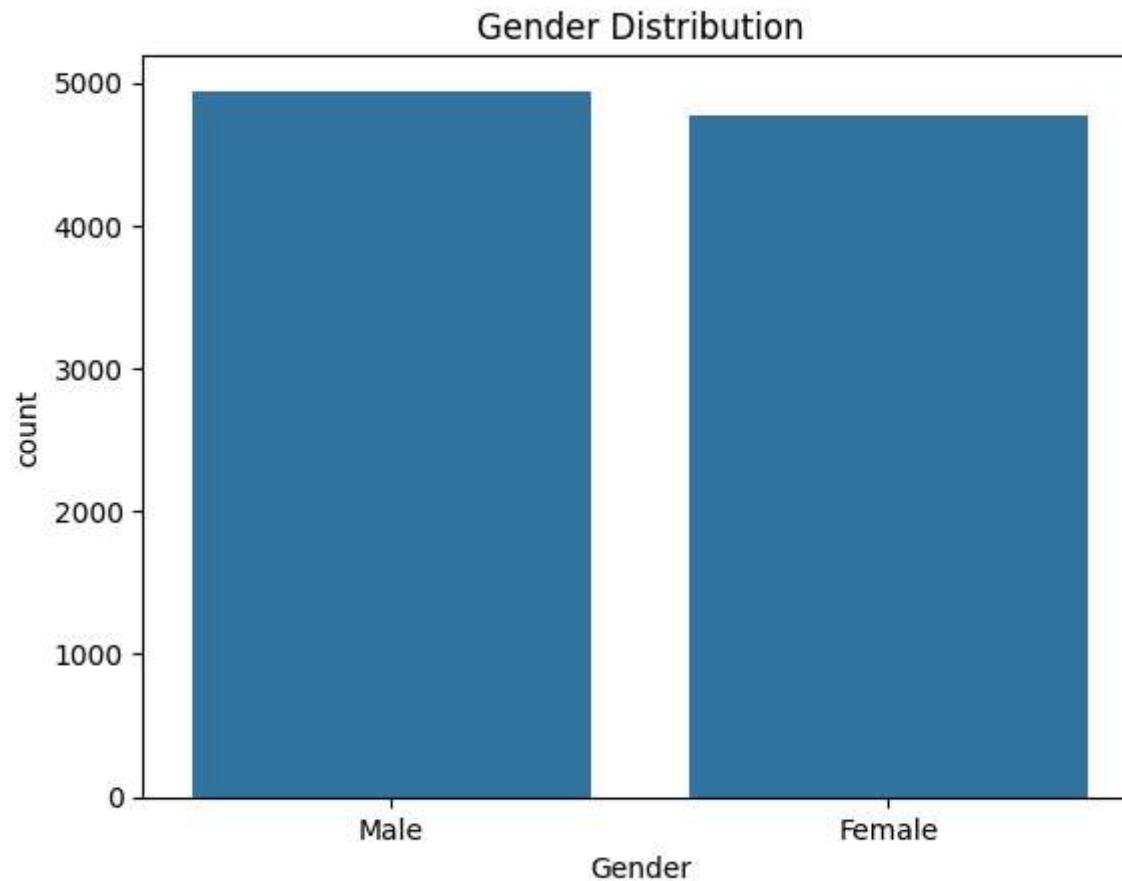
Age Distribution

```
In [11]: sns.histplot(df["Age"], kde=True, bins=20)  
plt.title("Age Distribution")  
plt.xlabel("Age")  
plt.show()
```



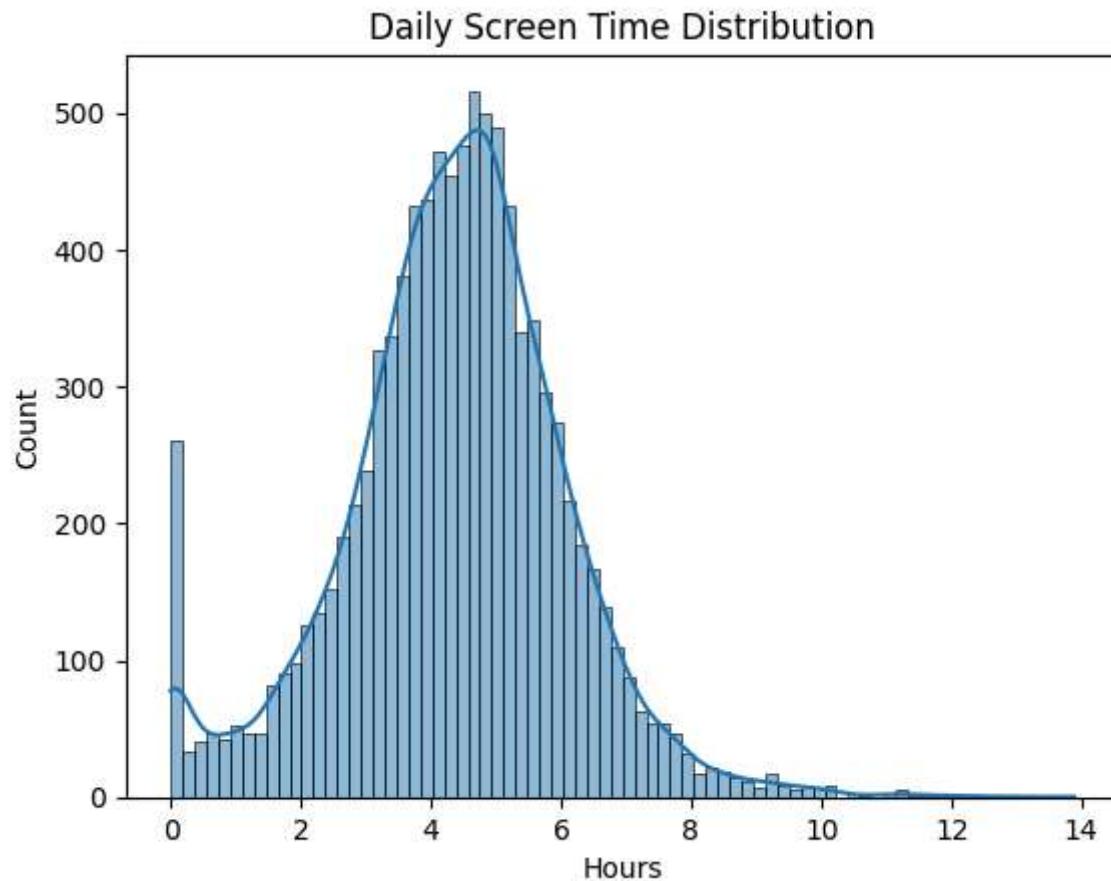
Gender Distribution

```
In [12]: sns.countplot(x="Gender", data=df)
plt.title("Gender Distribution")
plt.show()
```



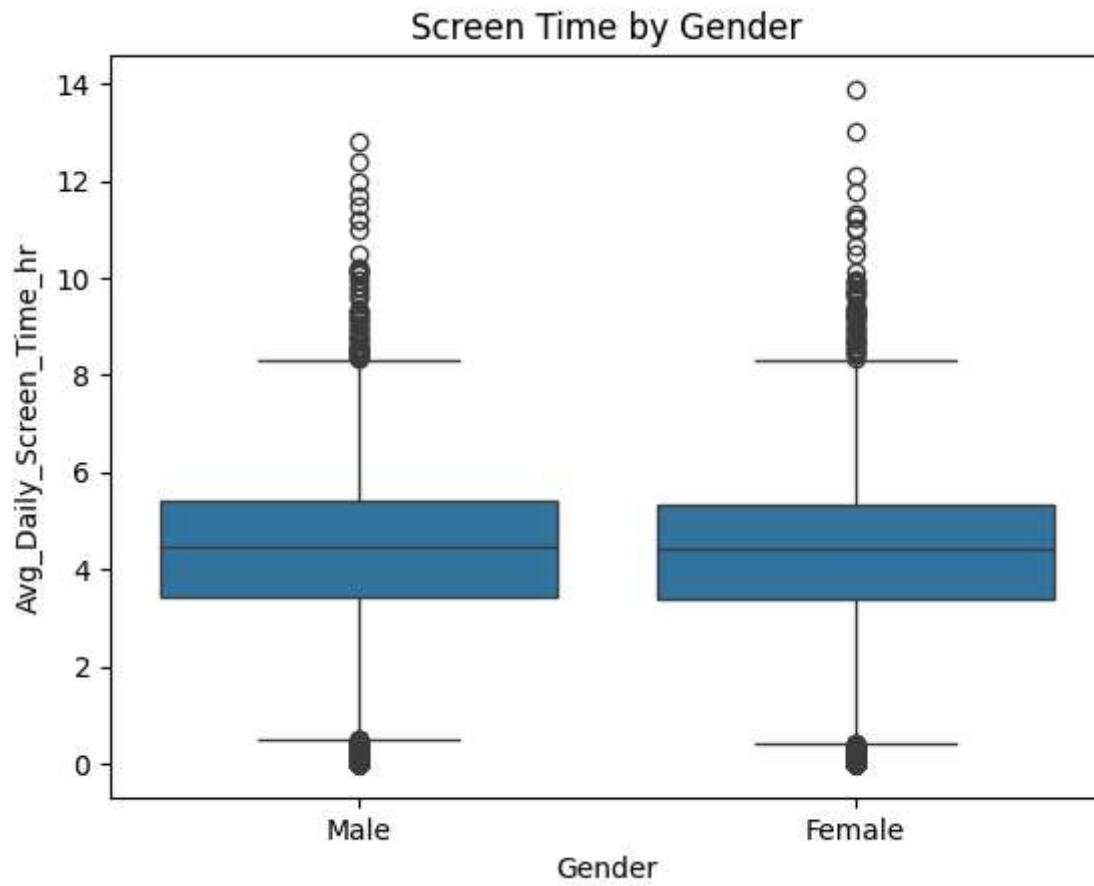
Avg Daily Screen Time

```
In [13]: sns.histplot(df["Avg_Daily_Screen_Time_hr"], kde=True)
plt.title("Daily Screen Time Distribution")
plt.xlabel("Hours")
plt.show()
```



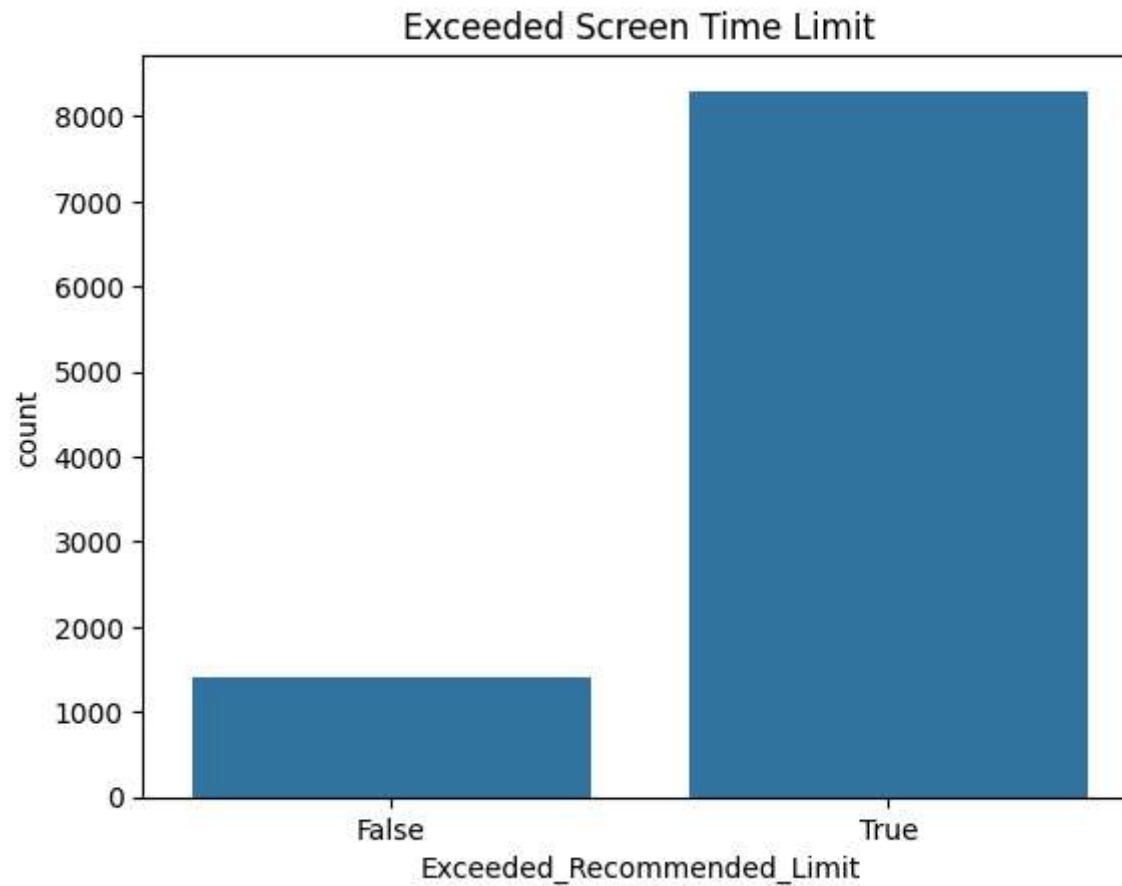
Screen Time by Gender

```
In [14]: sns.boxplot(x="Gender", y="Avg_Daily_Screen_Time_hr", data=df)
plt.title("Screen Time by Gender")
plt.show()
```



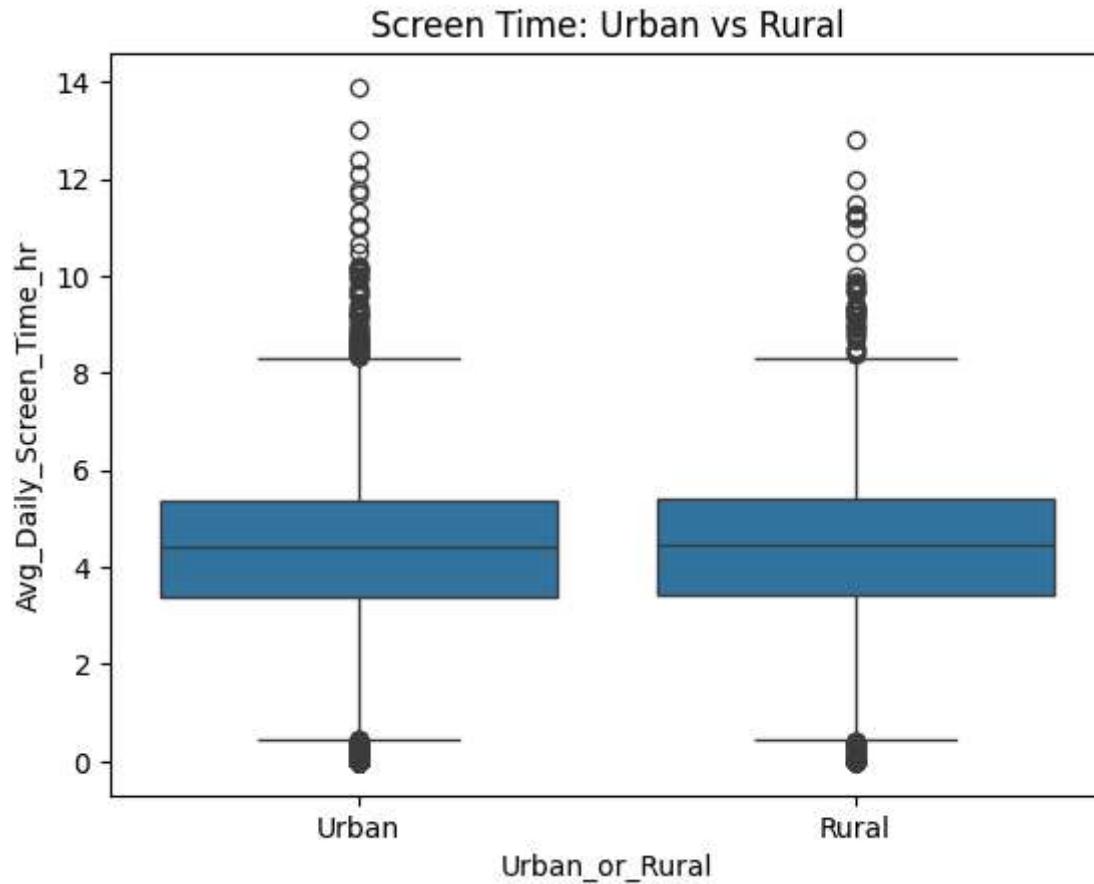
Exceeded Recommended Limit

```
In [15]: sns.countplot(x="Exceeded_Recommended_Limit", data=df)
plt.title("Exceeded Screen Time Limit")
plt.show()
```



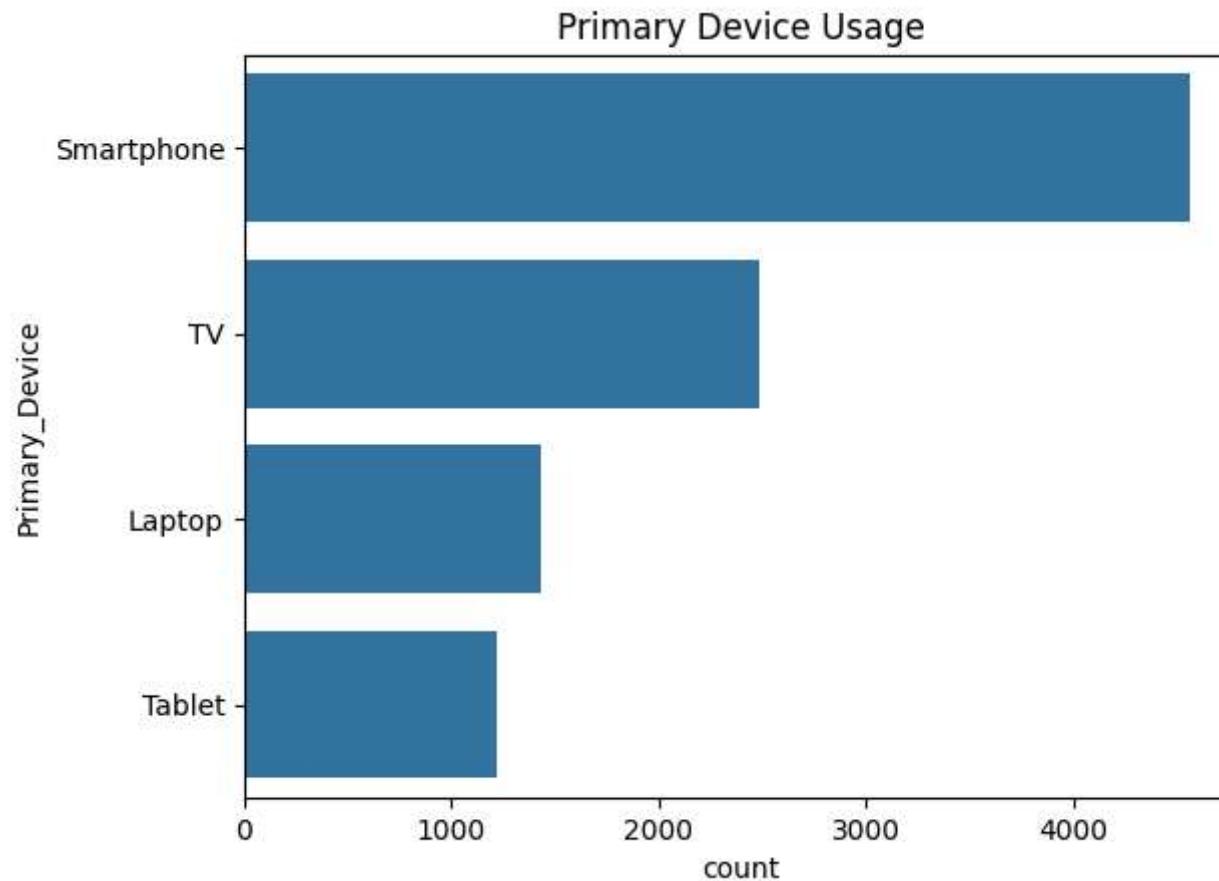
Screen Time by Urban vs Rural

```
In [16]: sns.boxplot(x="Urban_or_Rural", y="Avg_Daily_Screen_Time_hr", data=df)
plt.title("Screen Time: Urban vs Rural")
plt.show()
```



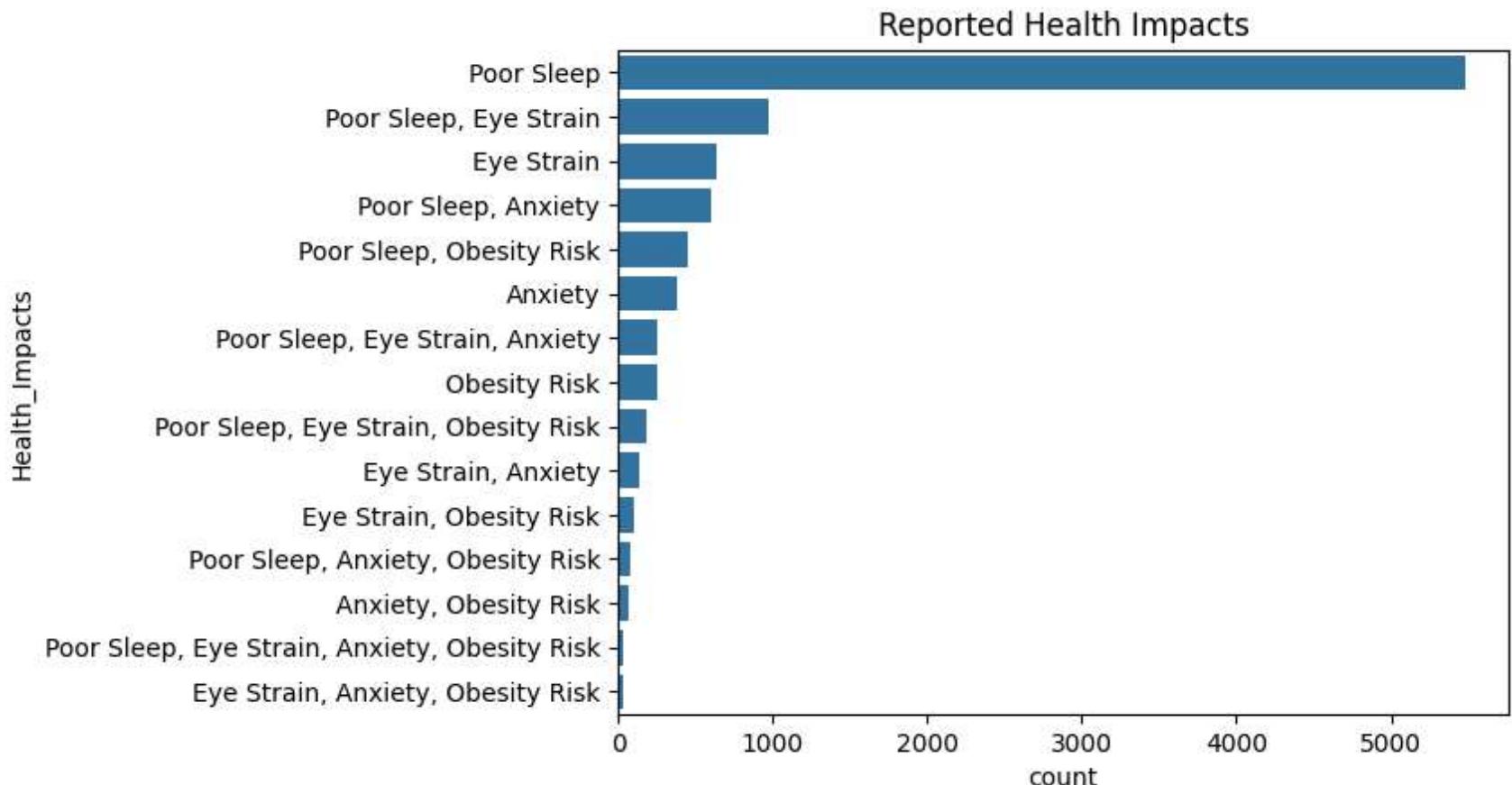
Health Impacts

```
In [17]: sns.countplot(y="Primary_Device", data=df, order=df["Primary_Device"].value_counts().index)
plt.title("Primary Device Usage")
plt.show()
```



Health Impacts vs Screen Time

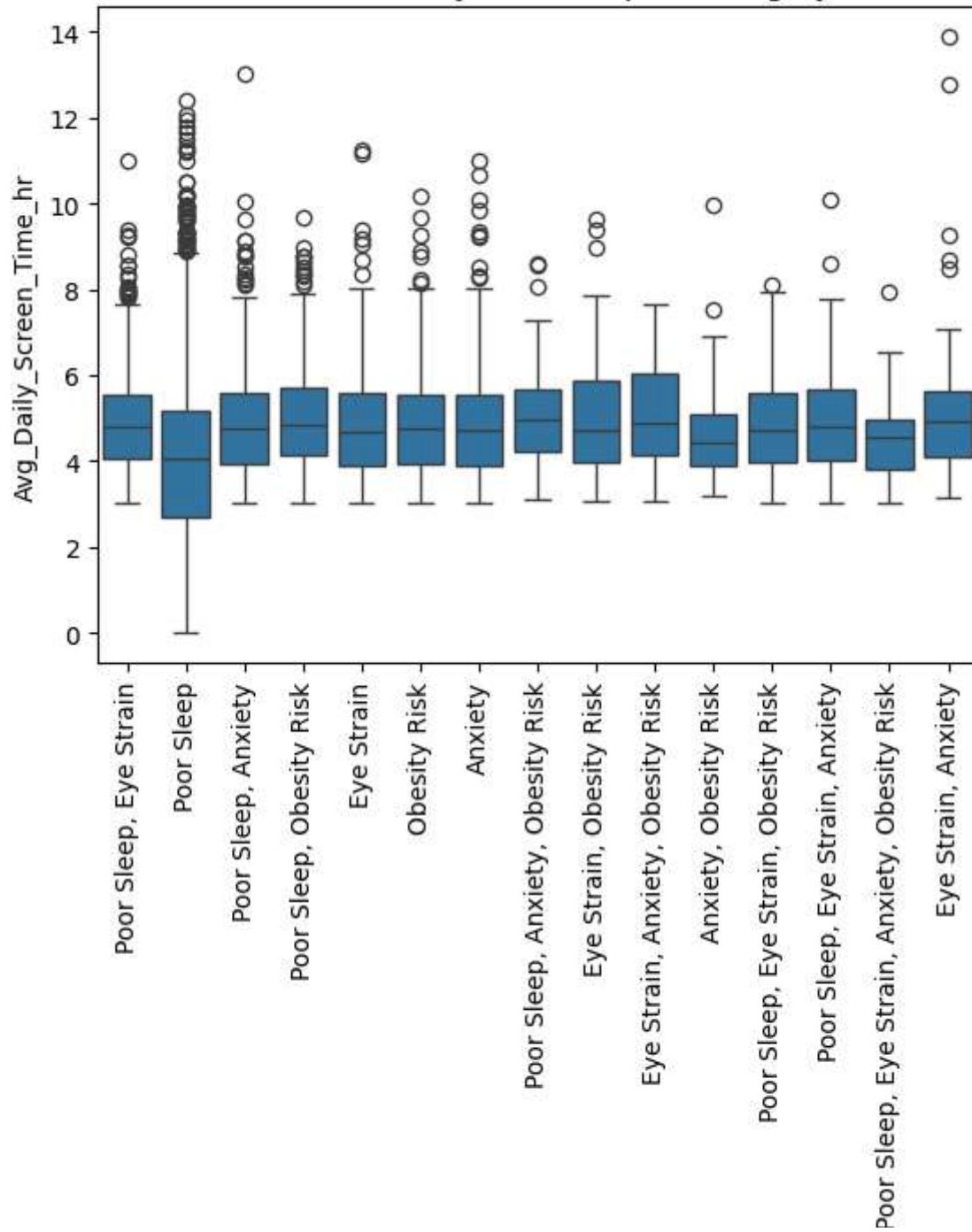
```
In [18]: sns.countplot(y="Health_Impacts", data=df, order=df["Health_Impacts"].value_counts().index)
plt.title("Reported Health Impacts")
plt.show()
```



Education to Recreation Ratio Distribution

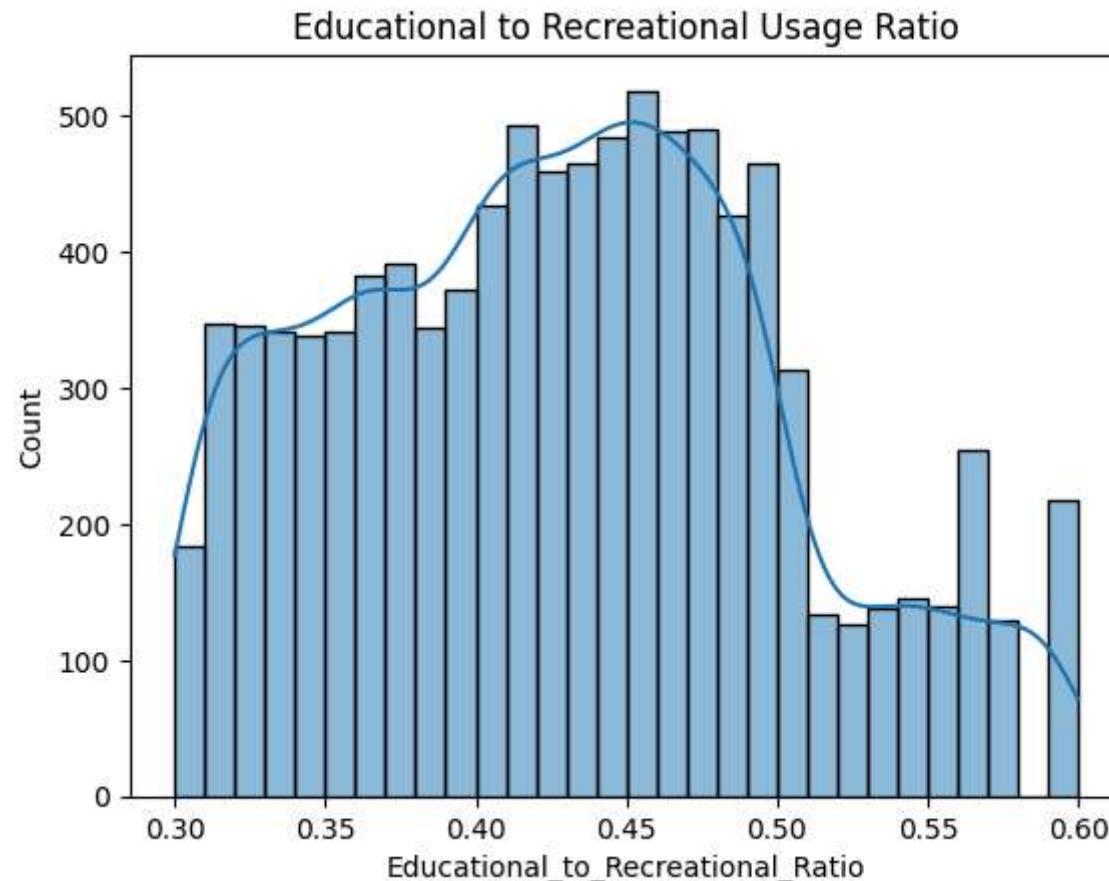
```
In [19]: sns.boxplot(x="Health_Impacts", y="Avg_Daily_Screen_Time_hr", data=df)
plt.title("Screen Time by Health Impact Category")
plt.xticks(rotation=90)
plt.show()
```

Screen Time by Health Impact Category



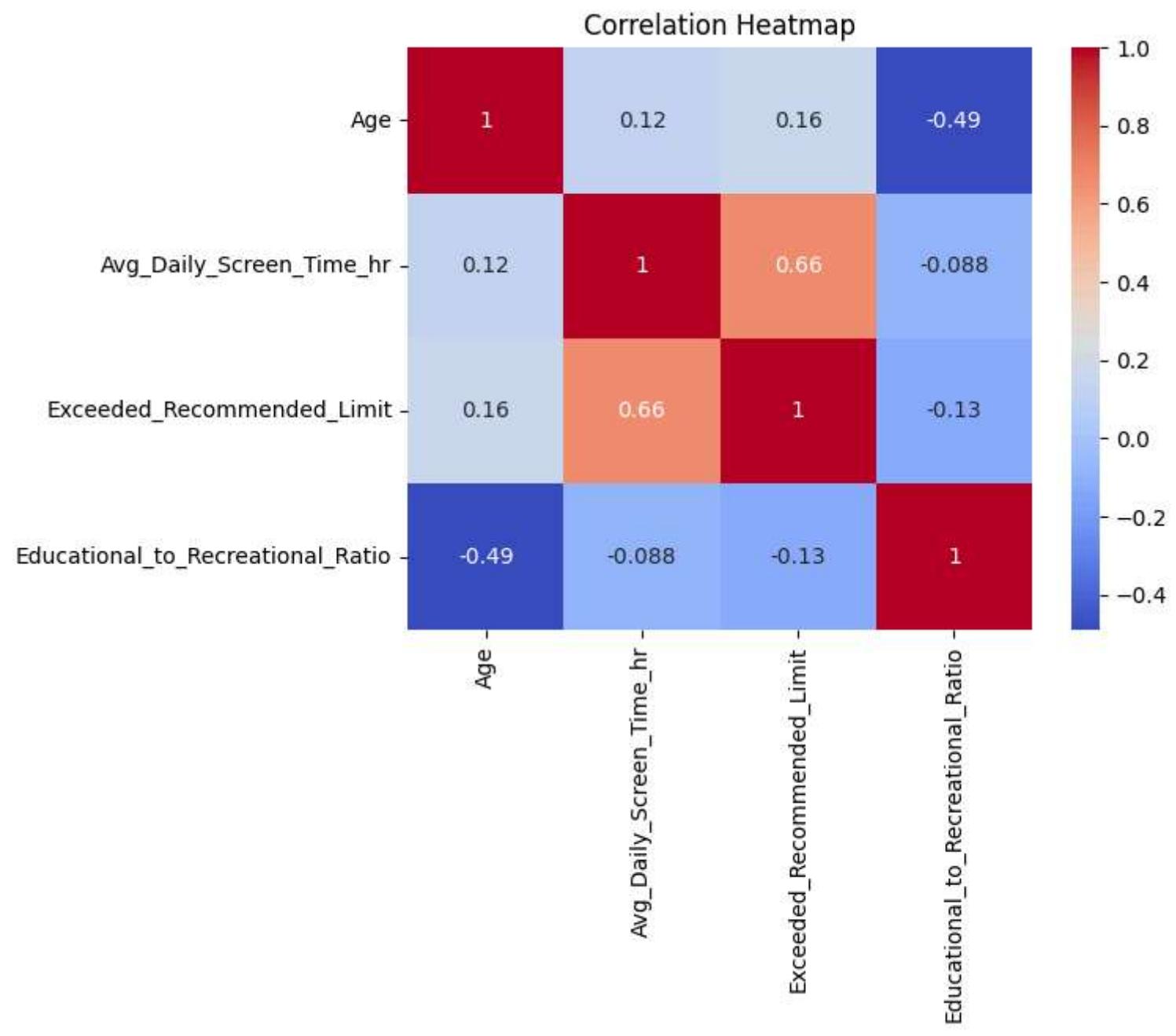
Education to Recreation Ratio Distribution

```
In [20]: sns.histplot(df["Educational_to_Recreational_Ratio"], kde=True)  
plt.title("Educational to Recreational Usage Ratio")  
plt.show()
```



Correlation Analysis (Numerical Columns)

```
In [21]: correlation = df.corr(numeric_only=True)
sns.heatmap(correlation, annot=True, cmap="coolwarm")
plt.title("Correlation Heatmap")
plt.show()
```



Age Grouping

In [22]:

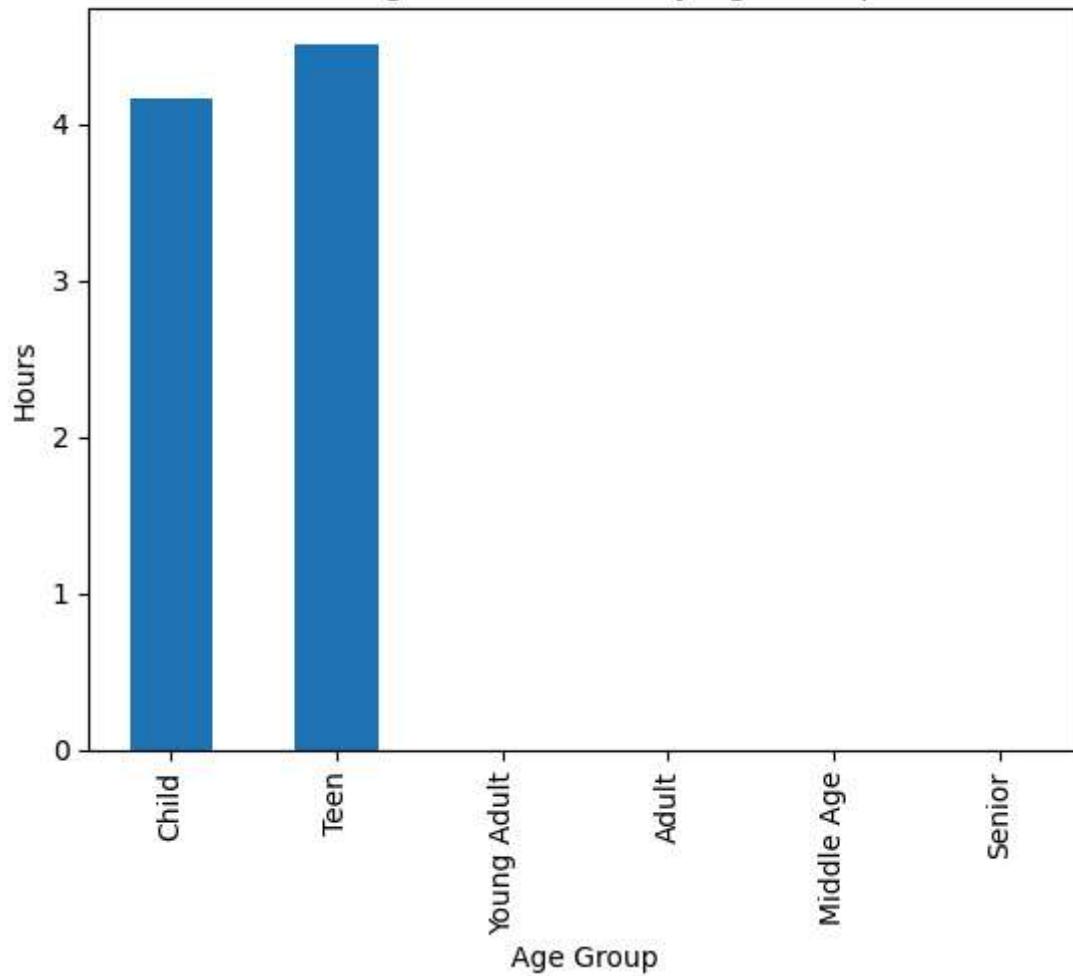
```
# Define age bins
bins = [0, 12, 18, 25, 40, 60, 100]
labels = ['Child', 'Teen', 'Young Adult', 'Adult', 'Middle Age', 'Senior']
df['Age_Group'] = pd.cut(df['Age'], bins=bins, labels=labels)

# Average screen time per age group
df.groupby('Age_Group')['Avg_Daily_Screen_Time_hr'].mean().plot(kind='bar')
plt.title("Average Screen Time by Age Group")
plt.ylabel("Hours")
plt.xlabel("Age Group")
plt.show()
```

C:\Users\hi\AppData\Local\Temp\ipykernel_16264\982581758.py:7: FutureWarning: The default of observed=False is deprecated and will be changed to True in a future version of pandas. Pass observed=False to retain current behavior or observed=True to adopt the future default and silence this warning.

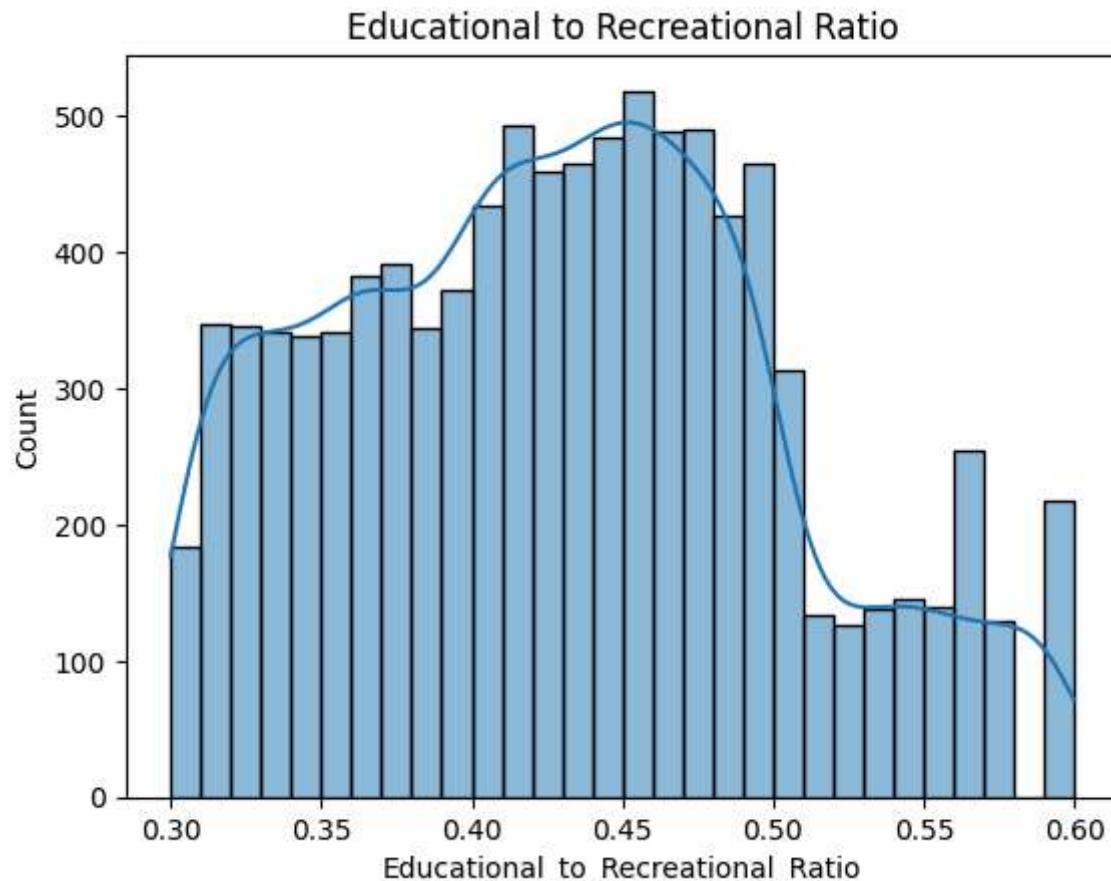
```
df.groupby('Age_Group')['Avg_Daily_Screen_Time_hr'].mean().plot(kind='bar')
```

Average Screen Time by Age Group



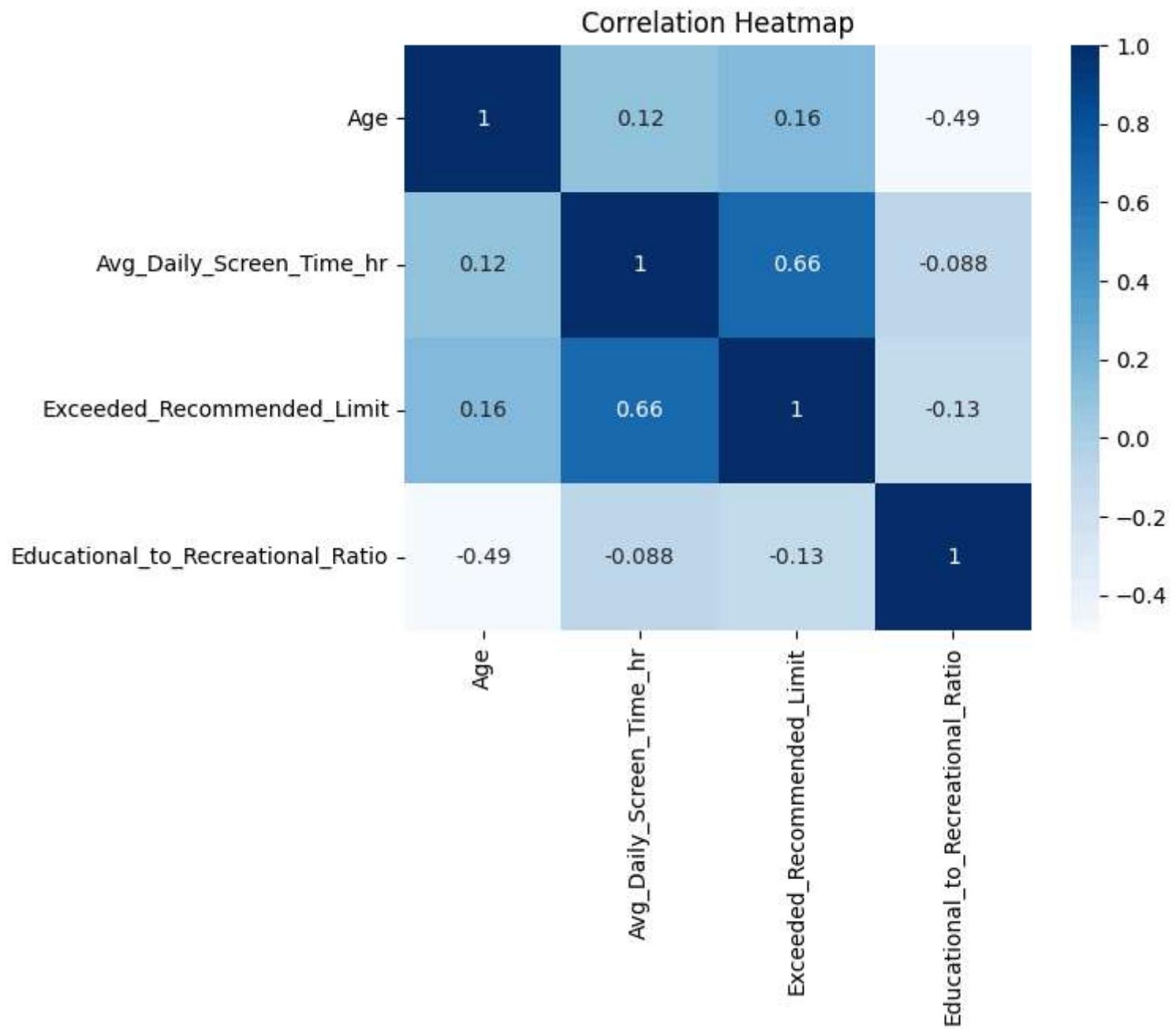
Educational to Recreational Ratio Analysis

```
In [23]: sns.histplot(df['Educational_to_Recreational_Ratio'], kde=True)
plt.title("Educational to Recreational Ratio")
plt.show()
```



In []:

```
In [24]: correlation = df.corr(numeric_only=True)
sns.heatmap(correlation, annot=True, cmap='Blues')
plt.title("Correlation Heatmap")
plt.show()
```



```
In [26]: pd.crosstab(df['Primary_Device'], df['Urban_or_Rural'])
```

Out[26]: `Urban_or_Rural` `Rural` `Urban`

Primary_Device	Urban	Rural
Laptop	404	1029
Smartphone	1320	3248
TV	745	1742
Tablet	392	832

Final Conclusion

The analysis of screentime data for Indian children in 2025 reveals a clear and growing dependence on digital devices across all age groups. Key findings indicate that:

- **Screentime increases with age**, with teenagers showing significantly higher daily usage than younger children.
- **Educational purposes** remain a major reason for screentime, but **entertainment and gaming** are rapidly catching up.
- Urban children generally exhibit **higher average screentime** than their rural counterparts.
- **Parental supervision** and **screen usage rules** strongly correlate with reduced daily screentime.
- Children with higher screentime report **reduced physical activity** and **shorter sleep durations**, pointing to potential health concerns.

These insights suggest the need for **balanced digital engagement strategies** involving parents, educators, and policymakers. Promoting digital well-being and integrating screen-time guidelines into family and school routines will be crucial in ensuring healthy development in the digital age.