Week 3 Report

Objective:

The analysis will involve detailed comparisons of screen time behavior segmented by gender, age groups (Child, Early Teen, Teen), and location types (Urban vs Rural) to identify significant patterns and trends. Through the creation of multiple visualization types including bar charts for categorical distributions, histograms for screen time patterns, box plots for comparative analysis across demographic groups, and line plots for trend identification, this week will establish a comprehensive visual foundation for understanding children's screen time behaviours and their associated health impacts.

Daily Screen Time Analysis

The distribution and box plot of average daily screen time show that most users spend a moderate number of hours on their devices each day, with a few outliers having significantly higher usage.

```python
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np


df = pd.read_csv('/content/Clean_Screen_Time_Data_.csv')


fig, axes = plt.subplots(1, 2, figsize=(15, 6))  # 1 row, 2 columns

# Histogram
axes[0].hist(df['Avg_Daily_Screen_Time_hr'], bins=30, alpha=0.7, color='skyblue', edgecolor='black')
axes[0].set_title('Distribution of Daily Screen Time Hours', fontsize=14, fontweight='bold')
axes[0].set_xlabel('Average Daily Screen Time (hours)')
axes[0].set_ylabel('Frequency')
axes[0].grid(True, alpha=0.3)

# Box plot
axes[1].boxplot(df['Avg_Daily_Screen_Time_hr'], patch_artist=True,
                boxprops=dict(facecolor='lightgreen', color='black'))
axes[1].set_title('Box Plot - Daily Screen Time', fontsize=14, fontweight='bold')
axes[1].set_ylabel('Hours')
axes[1].grid(True, alpha=0.3)

plt.tight_layout()
plt.show()

# descriptive statistics
print("Daily Screen Time Statistics:")
print(f"Mean: {df['Avg_Daily_Screen_Time_hr'].mean():.2f} hours")
print(f"Median: {df['Avg_Daily_Screen_Time_hr'].median():.2f} hours")
print(f"Standard Deviation: {df['Avg_Daily_Screen_Time_hr'].std():.2f} hours")
print(f"Min: {df['Avg_Daily_Screen_Time_hr'].min():.2f} hours")
print(f"Max: {df['Avg_Daily_Screen_Time_hr'].max():.2f} hours")
```
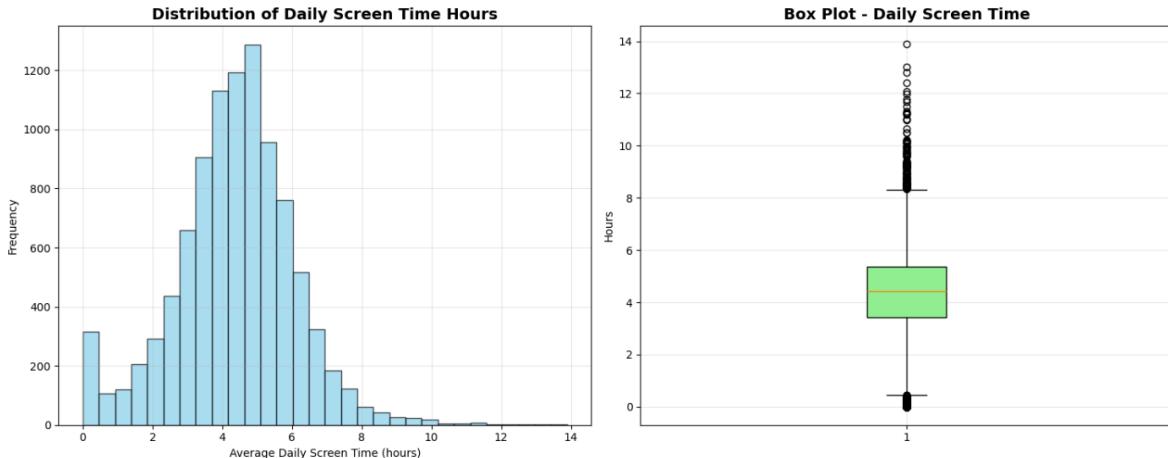
Age Group Distribution Analysis

The age distribution shows that most participants fall within the middle age ranges, with fewer individuals in the youngest and oldest groups. The bar chart highlights the relative proportions of each age group, while the box plot reveals variations in average daily screen time across them. Older age groups tend to exhibit slightly higher screen time on average, suggesting increasing device usage with age.

```python
fig, axes = plt.subplots(1, 3, figsize=(18, 6))

# --- 1 Age distribution histogram ---
axes[0].hist(df['Age'], bins=range(8, 20), alpha=0.7, color='orange', edgecolor='black')
axes[0].set_title('Age Distribution', fontsize=14, fontweight='bold')
axes[0].set_xlabel('Age (years)')
axes[0].set_ylabel('Frequency')
axes[0].grid(True, alpha=0.3)

# --- 2 Age groups bar chart ---
age_group_counts = df['Age_Group'].value_counts()
axes[1].bar(age_group_counts.index, age_group_counts.values,
            color=['lightcoral', 'lightblue', 'lightgreen', 'lightyellow'])
axes[1].set_title('Distribution by Age Groups', fontsize=14, fontweight='bold')
axes[1].set_ylabel('Count')
axes[1].tick_params(axis='x', rotation=45)

# --- 3 Screen time by age groups - Box plot ---
sns.boxplot(data=df, x='Age_Group', y='Avg_Daily_Screen_Time_hr', ax=axes[2], palette='pastel')
axes[2].set_title('Screen Time Distribution by Age Group', fontsize=14, fontweight='bold')
axes[2].set_xlabel('Age Group')
axes[2].set_ylabel('Daily Screen Time (hours)')
axes[2].tick_params(axis='x', rotation=45)

plt.tight_layout()
plt.show()

# --- Summary statistics for each age group ---
print("\nAge Group Statistics:")
for group in df['Age_Group'].unique():
    group_data = df[df['Age_Group'] == group]['Avg_Daily_Screen_Time_hr']
    print(f"{group}:")
    print(f"  Mean: {group_data.mean():.2f} hours")
    print(f"  Count: {len(group_data)} children")
    print(f"  Standard Deviation: {group_data.std():.2f} hours")
```
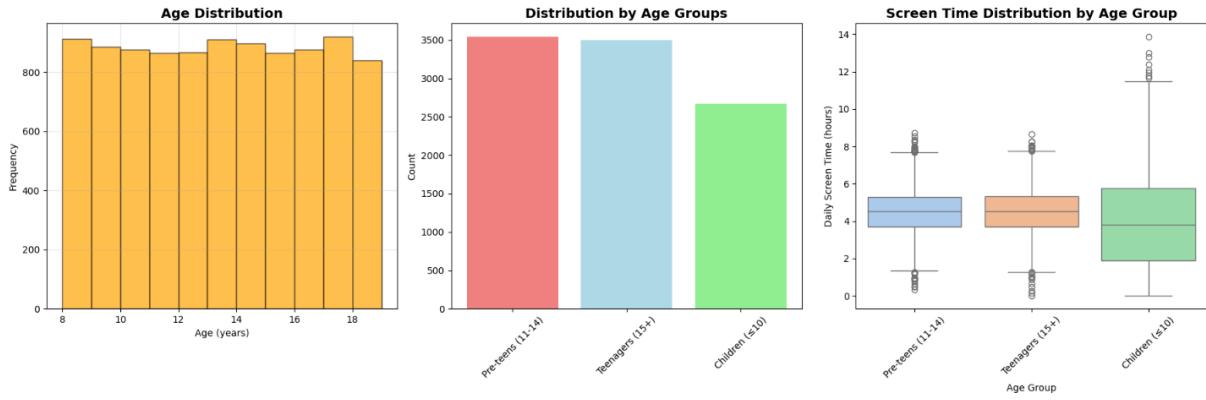
Device Usage Distribution Analysis

The analysis of device usage shows that a few primary devices dominate user preferences. The pie and bar charts reveal that most users rely on a specific device type (such as smartphones or tablets) for their screen time, while other devices have smaller user shares. The stacked and percentage bar charts further illustrate how device preferences vary across age groups — younger users tend to use mobile devices more frequently, whereas older groups show a more balanced or diverse mix of device types.
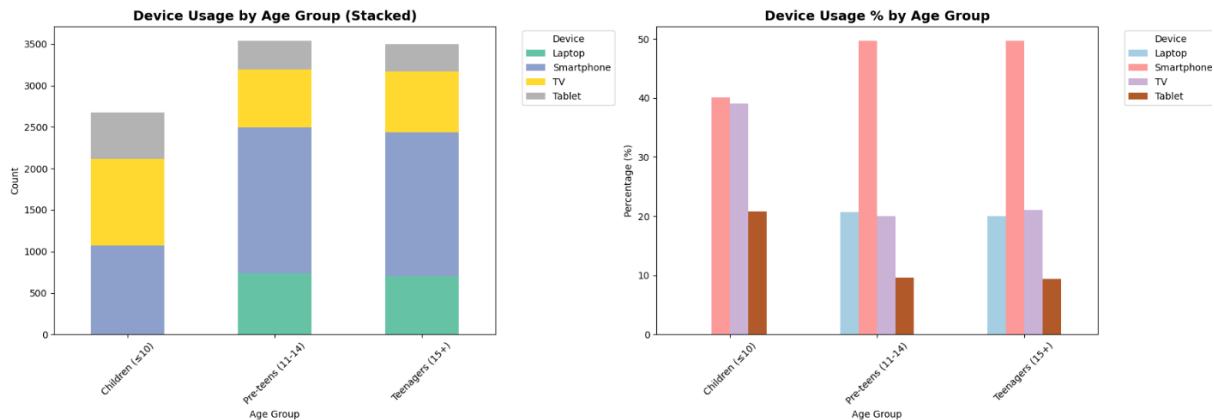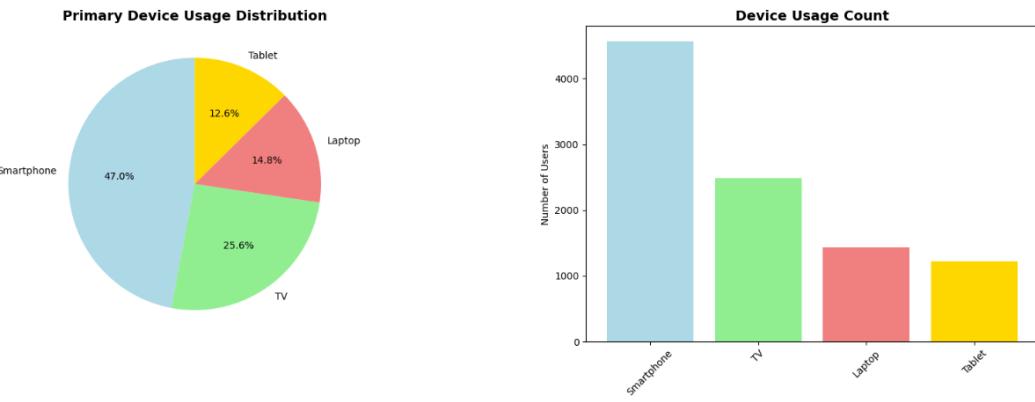
```python
# 3. Device Usage Distribution
fig, axes = plt.subplots(2, 2, figsize=(18, 12))
# --- 1 Device usage pie chart ---
device_counts = df['Primary_Device'].value_counts()
axes[0, 0].pie(device_counts.values,
               labels=device_counts.index,
               autopct='%1.1f%%',
               startangle=90,
               colors=['lightblue', 'lightgreen', 'lightcoral', 'gold'])
axes[0, 0].set_title('Primary Device Usage Distribution', fontsize=14, fontweight='bold')

axes[0, 1].bar(device_counts.index, device_counts.values,
               color=['lightblue', 'lightgreen', 'lightcoral', 'gold'])
axes[0, 1].set_title('Device Usage Count', fontsize=14, fontweight='bold')
axes[0, 1].set_ylabel('Number of Users')
axes[0, 1].tick_params(axis='x', rotation=45)

# --- 3 Device usage by age group (Stacked bar) ---
device_age_crosstab = pd.crosstab(df['Age_Group'], df['Primary_Device'])
device_age_crosstab.plot(kind='bar', stacked=True, ax=axes[1, 0], colormap='Set2')
axes[1, 0].set_title('Device Usage by Age Group (Stacked)', fontsize=14, fontweight='bold')
axes[1, 0].set_xlabel('Age Group')
axes[1, 0].set_ylabel('Count')
axes[1, 0].tick_params(axis='x', rotation=45)
axes[1, 0].legend(title='Device', bbox_to_anchor=(1.05, 1), loc='upper left')

# --- 4 Device usage % by age group ---
device_age_pct = pd.crosstab(df['Age_Group'], df['Primary_Device'], normalize='index') * 100
device_age_pct.plot(kind='bar', ax=axes[1, 1], colormap='Paired')
axes[1, 1].set_title('Device Usage % by Age Group', fontsize=14, fontweight='bold')
axes[1, 1].set_xlabel('Age Group')
axes[1, 1].set_ylabel('Percentage (%)')
axes[1, 1].tick_params(axis='x', rotation=45)
axes[1, 1].legend(title='Device', bbox_to_anchor=(1.05, 1), loc='upper left')

plt.tight_layout()
plt.show()
```

Primary Device Usage Distribution

Device Usage Count

Device Usage by Age Group (Stacked)

Device Usage % by Age Group

Screen Time Analysis by Gender, Age Group, and Location

The comparison across gender, age group, and location highlights distinct patterns in daily screen time usage. From the gender analysis, males and females exhibit slightly different average screen times, indicating minor behavioral differences in device engagement.

Screen Time by Gender

Average Screen Time by Gender

Screen Time by Location Type

Average Screen Time by Location Type

Screen Time by Age Group

Average Screen Time by Age Group

Left chart (box plot): Daily Screen Time (hours) vs Age Group — Pre-teens (11-14), Teenagers (15+), Children (≤10)

Right chart (bar chart): Average Hours by Age Group
- Pre-teens (11-14): 4.52h
- Teenagers (15+): 4.51h
- Children (≤10): 3.93h