User Engagement and App Usage Patterns Across Different Demographics

Mel Christian Aniban

John Rey Ortigas

BSCS 3-B

December 27, 2024

EXECUTIVE SUMMARY

The analysis explores a dataset containing detailed information on mobile app usage, daily screen time, and lifestyle patterns across different demographics, including age groups, genders, and geographic locations. The dataset provides insights into how app usage habits, screen time, and age-specific preferences influence digital engagement. This study offers a foundational basis for developing strategies to manage screen time, optimize app engagement, and promote healthier digital habits while contributing to a broader understanding of the interplay between technology use and lifestyle.

Key findings from the analysis reveal significant insights into app usage patterns and screen time behavior. Younger age groups, particularly teens and adolescents, exhibit the highest daily screen time, averaging nearly 10 hours per day. Gender-specific trends indicate that both males and females dedicate similar time to mobile apps with slight variations in category preference. Age-group analysis highlights that teens and adolescents prioritize social media and gaming apps, while middle-aged adults lean more towards productivity tools. Geographical differences also play a role, with locations like Phoenix and Los Angeles displaying slightly higher app engagement compared to other cities.

The findings underscore the importance of balancing screen time and app usage to promote overall well-being. Excessive screen time among younger demographics suggests a need for targeted interventions to encourage healthier digital practices. Strategies could include promoting digital literacy, setting app usage limits, and enhancing awareness about the impact of prolonged screen exposure on mental health. Additionally, the study provides a valuable resource for app developers, educators, and policymakers to tailor content and campaigns that cater to the diverse preferences of age groups and genders.The insights derived can inform future research, aiding stakeholders such as educators, healthcare professionals, and policymakers in understanding and addressing the factors influencing digital engagement. By leveraging these findings, targeted initiatives can be designed to improve digital well-being and promote balanced lifestyles across diverse populations.

INTRODUCTION

Objective

- To analyze daily screen time across different age groups and identify trends.
- To explore app usage patterns based on gender and examine differences in time spent on various app categories.
- To evaluate app usage across different age groups focusing on social media, gaming, and productivity apps.
- To assess app usage by location to uncover potential geographic trends in mobile behavior.

Data Source

The Smartphone Usage and Behavioral Dataset is sourced from Kaggle, a prominent data science platform. It contains detailed information about smartphone app usage, daily screen time, and behavioral patterns across various demographics, including age, gender, and geographic location. The dataset was designed to provide insights into how different populations interact with their smartphones, offering a robust foundation for analyzing trends in digital behavior and lifestyle.

Relevance

This dataset is a valuable resource for understanding how smartphone usage and screen time impact individual and societal well-being. It is particularly relevant for app developers, educators, healthcare professionals, and policymakers looking to promote healthier digital habits, optimize app engagement, and mitigate the effects of excessive screen time. The insights derived can inform the development of targeted interventions, such as digital literacy programs, screen time management strategies, and campaigns to address the mental and physical health effects of prolonged smartphone use. Moreover, the dataset supports future research efforts, contributing to a broader understanding of the relationship between technology use and lifestyle across diverse populations.

DATASET OVERVIEW

Dataset Name: Smartphone Usage and Behavioral Dataset

Source: [Smartphone Usage and Behavioral Dataset]

Description: This dataset contains the daily mobile usage patterns of 1,000 users, covering aspects such as screen time, app usage, user engagement across different app categories and user location.

Rows and Columns:

- **Rows:** 1,000 rows (each representing each user's smartphone usage and behavioral data identified by unique User_id).
- **Columns:** 12 columns (*e.g., User_id, Age, Gender, Location, Total mobile app usage hours, Daily screen time hours, Number of apps used, Social media usage hours, Productivity app usage hours, Gaming app usage hours, Other activities on mobile, and Age group*).

Columns and Description:

Column name	**Description**	**Example**
User_id	A unique identifier for each individual in the dataset.	3
Age	The age of the user, expressed in years.	32
Gender	The user's gender, typically categorized as Male or Female.	Female
Total_mobile_app_usage_hours	Total hours spent using all mobile apps daily.	9.12
Daily_screen_time_hours	Total hours the user spends on their mobile device daily.	9.12
Number_of_apps_used	Count of distinct apps the user engages with daily.	11
Social_media_usage_hours	Hours spent using social media apps daily.	4.58

Productivity_app_usage_hours	Hours spent on productivity apps such as work-related or educational tools daily.	1.71
Gaming_app_usage_hours	Hours spent on gaming apps daily.	2.83
Other_act_in_using_mobile	Time spent on other activities not classified as social media, productivity, or gaming.	0.00
Location	The city or region where the user resides.	Houston
Demographic	The categorization of users into age-based groups	Young Adults

METHODOLOGY

The methodology for this case study encompasses a structured series of steps aimed at preparing and analyzing the dataset. This process includes data cleaning, wrangling, and in-depth analysis, facilitating the creation of insightful visualizations and the derivation of meaningful conclusions.

1. **Data Acquisition**
    - Downloaded the dataset for analysis.
    - Imported the dataset (CSV format) along with the necessary libraries
    - required for processing and analysis.
2. **Data Cleaning**
    - Checked for missing values in all columns and identified duplicate entries to maintain data integrity.
    - Standardized column names for consistency by stripping spaces, capitalizing, and replacing spaces with underscores.
    - Computed the sum of individual app usage columns and renamed the column for total app usage.
    - Addressed discrepancies by ensuring the total app usage aligns with the sum of individual usage columns.
    - Resolved inconsistencies in total screen time.
    - Derived remaining time spent on mobile activities beyond categorized app usage.
    - Group Users into Age Groups
    - Drop Unnecessary Columns
3. **Data Analysis**
    - Analyzed average screen time by different Age groups by using **groupby()**
    - Analyzed average app usage in different categories of mobile apps by gender by using **groupby()**.

- Analyzed average app usage in different categories of mobile apps by Age groups by using **groupby()**.
- Analyzed average app usage in different categories of mobile apps usage by Location by using **groupby().**

4. **Data Visualization**
    - Created a bar chart to show daily screen time across different age groups.
    - Created a bar chart to show different categories of mobile apps usage by gender.
    - Created a bar chart to show different categories of mobile apps usage by Age groups **.**
    - Created a bar chart to show different categories of mobile apps usage by Location.

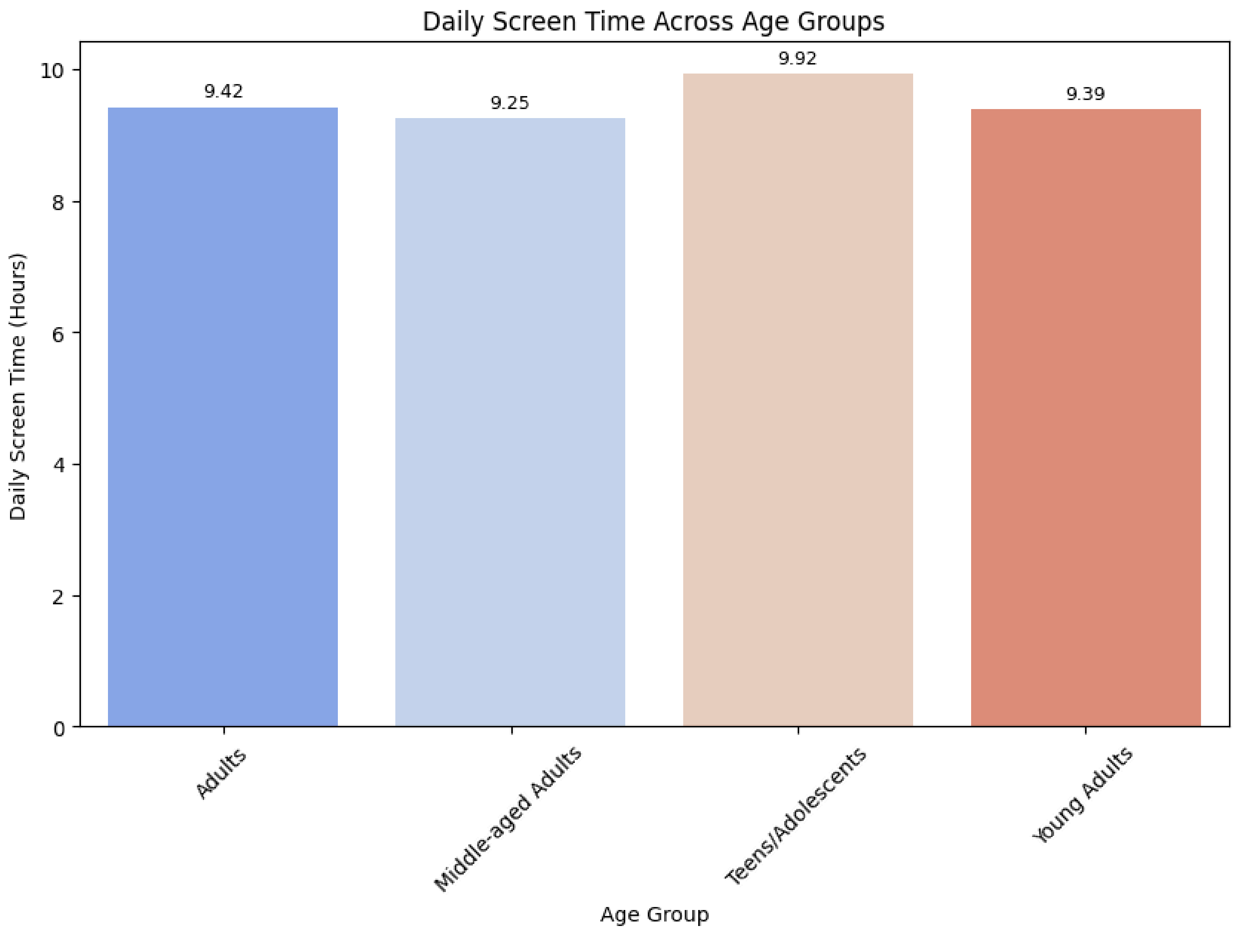
RESULT AND ANALYSIS

Key Findings

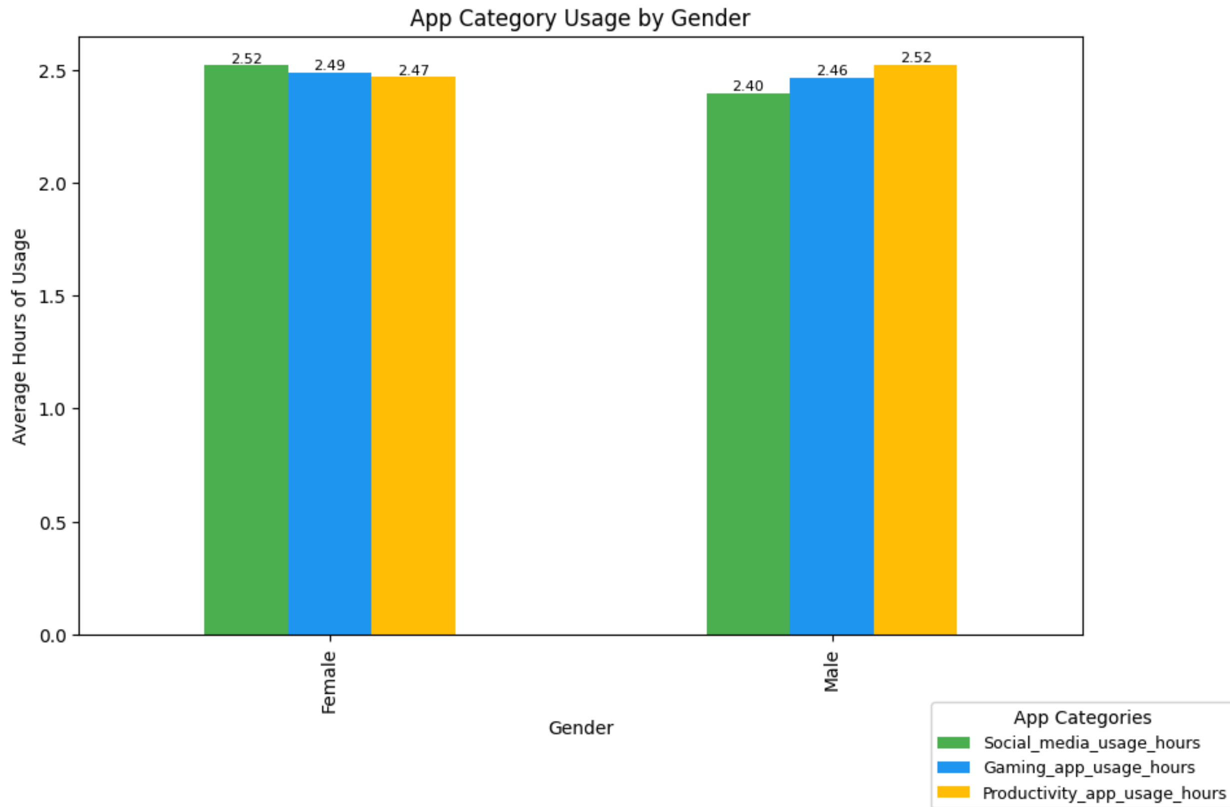
1. **Daily Screen Time:** Teens and adolescents have the highest daily screen time, while middle-aged adults spend the least time on screens.
2. **Gender-based App Usage:** Females predominantly use social media apps, while males prefer productivity apps.
3. **App Usage by Age Group:** Teens focus most on social media and gaming, whereas adults lead in productivity app usage.
4. **Geographic Trends:** Productivity app usage is highest in Phoenix and New York, social media usage peaks in Houston, and gaming dominates in Chicago.

Visualizations

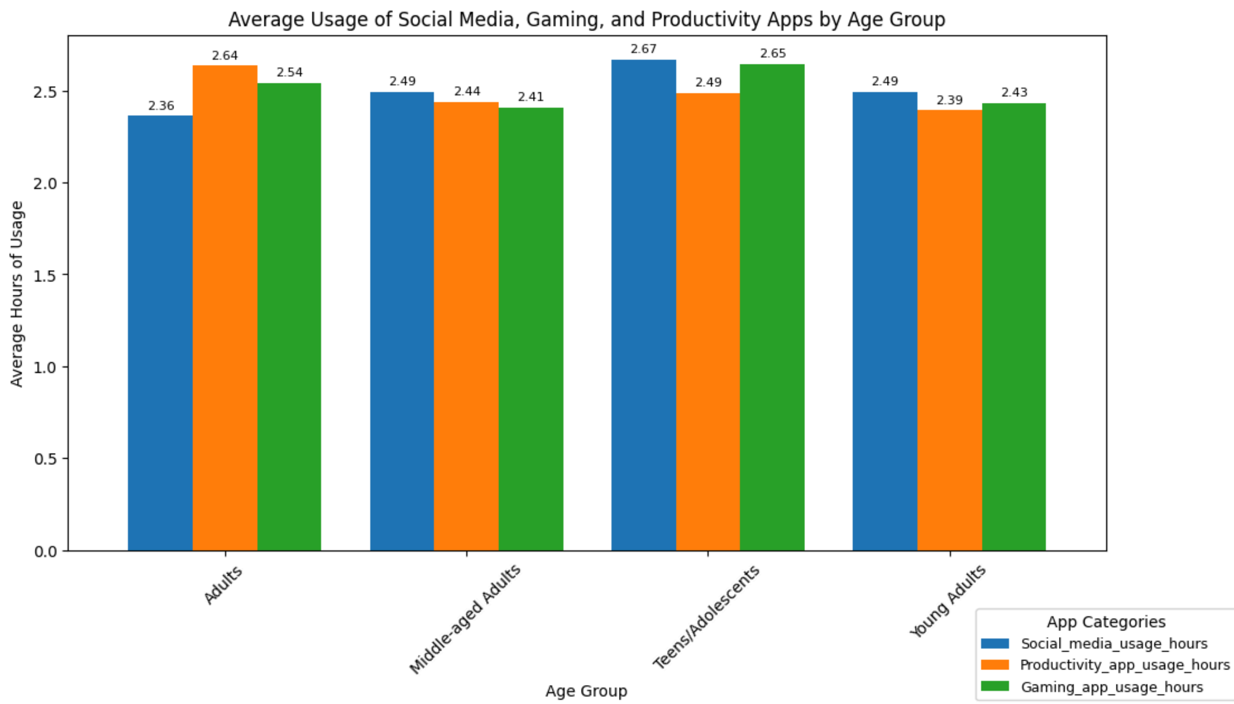
1. **Bar Graph:** Daily ScreenTime Across Age Groups



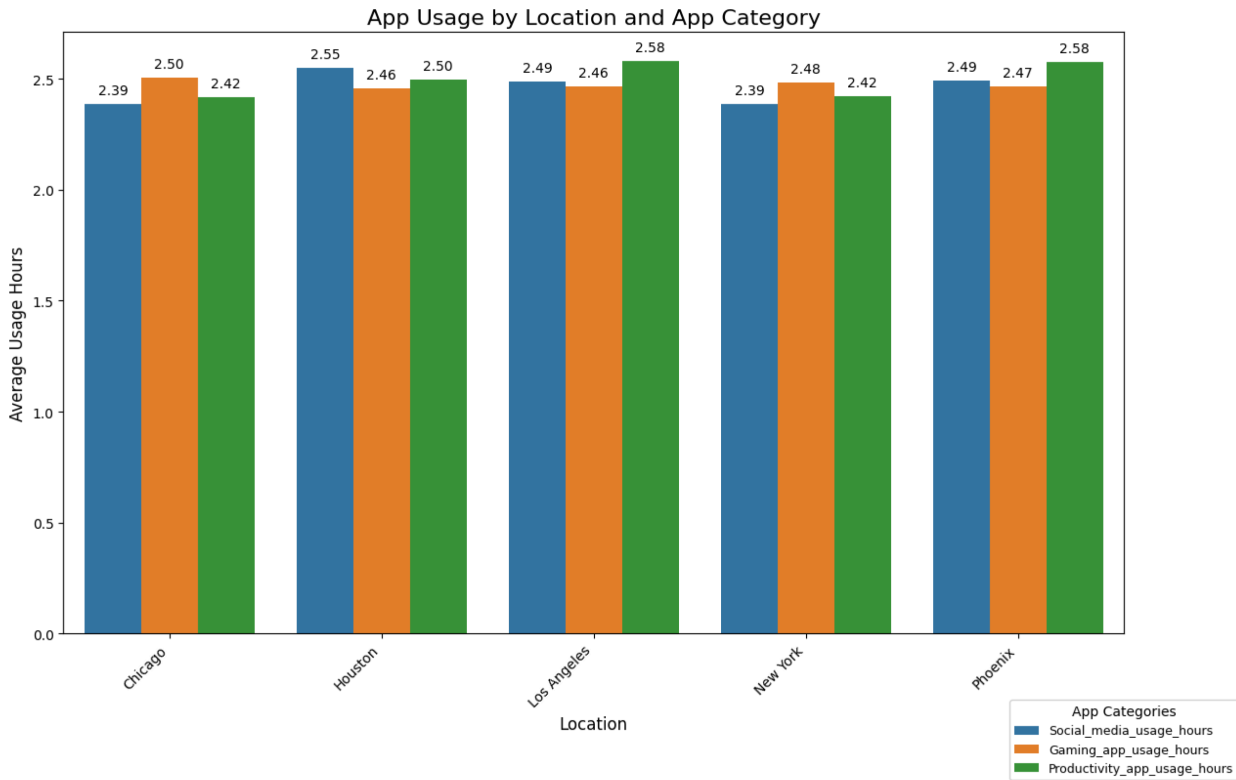
2. Multiple bar graph (Grouped bar graph): App Category Usage by Gender



3. **Multiple bar graph (Grouped bar graph)**: Average Usage of Social Media, Gaming, Productivity Apps by Age Groups



4. **Multiple bar graph (Grouped bar graph):**App Usage by Location And App Category



INSIGHTS AND RECOMMENDATIONS

Insights

- Teens and adolescents exhibit the highest daily screen time, suggesting they are the most engaged demographic in mobile activities.
- Middle-aged adults, with the lowest screen time, may prioritize non-digital activities or have less inclination towards mobile app engagement.
- Females gravitate towards social media apps, indicating their preference for communication, networking, and content sharing platforms.
- Males show higher usage of productivity apps, suggesting they use mobile devices more for work or task-oriented purposes.
- Teens dominate social media and gaming usage, reflecting their focus on entertainment and social connections.
- Adults prioritize productivity apps, indicating their inclination towards work-related and organizational tools.
- High productivity app usage in Phoenix and New York reflects a professional and task-focused user base.
- Social media's peak in Houston suggests it is a hub for digital communication and content consumption.
- Gaming's prominence in Chicago highlights its popularity as a source of entertainment in the region.

Recommendations

- Enhance social media and gaming apps with features that appeal to younger audiences, such as gamification, interactive content, and exclusive rewards.
- Introduce parental controls or screen time reminders to promote healthier usage habits.
- Develop mobile apps or campaigns that cater to practical needs, such as fitness trackers, finance apps, or lifestyle management tools, to encourage increased engagement.
- For females: Focus on improving social media apps with enhanced community-building features, personalized content recommendations, and visual storytelling tools.

- For males: Promote productivity apps by integrating advanced features like cross-device compatibility, time management tools, and task automation.
- In Phoenix and New York: Expand marketing and feature enhancements for productivity apps to cater to the professional user base.
- In Houston: Invest in social media advertising and partnerships, emphasizing local content and influencer-driven campaigns.
- In Chicago: Collaborate with gaming studios to offer region-specific content, tournaments, or exclusive rewards.
- Introduce cross-category app integrations, such as blending social media with productivity tools or gaming with social sharing features.
- Utilize the insights to design targeted campaigns leveraging demographics and location-based data to increase app engagement and user retention.

CONCLUSION

This case study involved a structured and systematic approach to data analysis, beginning with data acquisition and cleaning, followed by exploratory analysis and visualization to uncover meaningful insights. Key steps included identifying and resolving missing and duplicate values, standardizing data formats, and aggregating metrics to ensure data integrity. Analyses revealed significant trends in screen time and app usage patterns across age groups, genders, and locations. These findings underscore the importance of tailoring digital strategies to demographic and geographic factors. By combining rigorous methodology with visualization techniques, this study highlighted actionable insights, contributing to a deeper understanding of mobile usage behaviors.

Challenges

- Limited data collection especially for the age groups of Children and Seniors/Elderly.
- External factors like cultural norms or occupation types are not included in the dataset, which could potentially lead to inaccuracies in the insights.
- Needs additional knowledge of Python libraries to effectively implement and achieve the desired objectives.

Reflection

Creating this case study was both challenging and rewarding. I found documentation particularly difficult, as it required detailed tracking of each step to ensure clarity and reproducibility. Exploring Python libraries in depth also felt overwhelming due to the vast range of functionalities and nuances in their use. Additionally, selecting an appropriate dataset posed its own challenges, with datasets either being overly complex, too messy, or unsuitable in size. Despite these hurdles, the process enhanced my analytical and problem-solving skills, providing valuable experience in handling real-world data scenarios and reinforcing the importance of persistence and adaptability in data analysis projects.

REFERENCES

American Psychological Association. (2020). *Digital media use and mental health*. Retrieved from https://www.apa.org/news/press/releases/stress/2020/digital-media-use

Anderson, M., & Jiang, J. (2018). Teens, social media, and technology 2018. *Pew Research Center*. Retrieved from https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/

Twenge, J. M., & Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports*, 12, 271-283. https://doi.org/10.1016/j.pmedr.2018.10.003

World Health Organization. (2019). *Guidelines on physical activity, sedentary behaviour, and sleep for children under 5 years of age*. Geneva: World Health Organization.

APPENDIX

Appendix A. Code for Data Acquisition

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px

df = pd.read_csv('mobile_usage_behavioral_analysis.csv')
```

Appendix B. Code for Data Cleaning

```
print(df.isnull().sum())
print(df.duplicated().sum())
print(df.describe())
```

Appendix C. Output of Data Cleaning

```
Null Values
User_ID                     0
Age                         0
Gender                      0
Total_App_Usage_Hours       0
Daily_Screen_Time_Hours     0
Number_of_Apps_Used         0
Social_Media_Usage_Hours    0
Productivity_App_Usage_Hours    0
Gaming_App_Usage_Hours      0
Location                    0
dtype: int64
Duplicated Values
0
           User_ID          Age  Total_App_Usage_Hours  \
count  1000.000000  1000.000000            1000.000000
mean    500.500000    38.745000               6.405670
std     288.819436    12.186734               3.134855
min       1.000000    18.000000               1.000000
25%     250.750000    28.000000               3.590000
50%     500.500000    40.000000               6.455000
75%     750.250000    50.000000               9.122500
max    1000.000000    59.000000              11.970000
```

```
       Daily_Screen_Time_Hours  Number_of_Apps_Used
Social_Media_Usage_Hours  \
count            1000.000000         1000.00000            1000.000000
mean                7.696310          16.64700              2.456330
std              3.714187           7.61961              1.439525
min               1.010000           3.00000              0.000000
25%                4.530000          10.00000               1.200000
50%                7.880000          17.00000              2.445000
75%               10.910000          23.00000               3.672500
max              14.000000          29.00000               4.990000

       Productivity_App_Usage_Hours  Gaming_App_Usage_Hours
count                 1000.000000             1000.000000
mean                     2.495270               2.475410
std                   1.443392              1.450362
min                    0.000000              0.010000
25%                      1.282500               1.220000
50%                      2.435000               2.455000
75%                      3.710000               3.782500
max                      5.000000              5.000000
```

Appendix D. Code for Data Wrangling

D1.

```
df.rename(columns={'Total_app_usage_hours':'Total_mobile_app_usage_hours'},
inplace = True)
```

```
df.columns = df.columns.str.strip().str.capitalize().str.replace(' ', '_')
```

D2.

```
df['Usage_sum_check'] = (
   df['Social_media_usage_hours'] +
   df['Productivity_app_usage_hours'] +
   df['Gaming_app_usage_hours']
)
df['Discrepancies'] = df['Total_mobile_app_usage_hours'] != df['Usage_sum_check']

df['Total_mobile_app_usage_hours'] = df.apply(lambda row: row['Usage_sum_check']
if row['Discrepancies'] == True else row['Total_mobile_app_usage_hours'], axis=1)
```

D3.

```
df['Discrepancies_in_total_screen_time'] = df['Total_mobile_app_usage_hours'] >=
```

```
df['Daily_screen_time_hours']

df['Daily_screen_time_hours'] = df.apply(lambda row:
row['Total_mobile_app_usage_hours'] if row['Discrepancies_in_total_screen_time'] ==
True else row['Daily_screen_time_hours'], axis=1)

df['Other_act_in_using_mobile'] =
df['Daily_screen_time_hours']-df['Total_mobile_app_usage_hours']
```

D4.

```
df['Age_group'] = df.apply(lambda row: row['Total_mobile_app_usage_hours'] if
row['Discrepancies_in_total_screen_time'] == True else
row['Daily_screen_time_hours'], axis=1)

df['Age_group'] = df['Age'].apply(lambda age:
    'Teens/Adolescents' if 13 <= age <= 19 else
    'Young Adults' if 20 <= age <= 34 else
    'Adults' if 35 <= age <= 49 else
    'Middle-aged Adults' if 50 <= age <= 64 else
    'Seniors/Elderly' if age >= 65 else
    'Children')
```

D5.

```
df_rmv_col = df.drop(['Discrepancies',
'Discrepancies_in_total_screen_time','Usage_sum_check'], axis=1)
```

Appendix E. Output for Data Wrangling

```
User_id  Age  Gender  Total_mobile_app_usage_hours  \
0        1   56    Male                          7.38
1        2   46    Male                         11.52
2        3   32  Female                          9.12
3        4   25  Female                         11.18
4        5   38    Male                          7.28

   Daily_screen_time_hours  Number_of_apps_used  Social_media_usage_hours  \
0                     7.38                   24                      4.43
1                    13.79                   18                      4.67
2                     9.12                   11                      4.58
3                    11.18                   21                      3.18
4                    12.59                   14                      3.15
```

```
  Productivity_app_usage_hours  Gaming_app_usage_hours     Location  \
0                         0.55                    2.40  Los Angeles
1                         4.42                    2.43      Chicago
2                         1.71                    2.83      Houston
3                         3.42                    4.58      Phoenix
4                         0.13                    4.00     New York

  Other_act_in_using_mobile           Age_group
0                      0.00  Middle-aged Adults
1                      2.27              Adults
2                      0.00        Young Adults
3                      0.00        Young Adults
4                      5.31              Adults
           User_id          Age  Total_mobile_app_usage_hours  \
count  1000.000000  1000.000000                   1000.000000
mean    500.500000    38.745000                      7.427010
std     288.819436    12.186734                      2.466041
min       1.000000    18.000000                      0.730000
25%     250.750000    28.000000                      5.710000
50%     500.500000    40.000000                      7.400000
75%     750.250000    50.000000                      9.075000
max    1000.000000    59.000000                     14.240000

       Daily_screen_time_hours  Number_of_apps_used  Social_media_usage_hours  \
count              1000.000000           1000.00000               1000.000000
mean                  9.392400             16.64700                  2.456330
std                   2.553704              7.61961                  1.439525
min                   2.540000              3.00000                  0.000000
25%                   7.567500             10.00000                  1.200000
50%                   9.355000             17.00000                  2.445000
75%                  11.442500             23.00000                  3.672500
max                  14.240000             29.00000                  4.990000

       Productivity_app_usage_hours  Gaming_app_usage_hours  \
count                   1000.000000             1000.000000
mean                       2.495270                2.475410
std                        1.443392                1.450362
min                        0.000000                0.010000
25%                        1.282500                1.220000
50%                        2.435000                2.455000
75%                        3.710000                3.782500
max                        5.000000                5.000000

       Other_act_in_using_mobile
count                1000.000000
mean                    1.965390
```

```
std              2.589749
min              0.000000
25%              0.000000
50%              0.260000
75%              3.660000
max             11.550000
(1000, 12)
```

Appendix F. Code for Data Analysis

F1.

```
age_group_summary = df_rmv_col.groupby('Age_group',
as_index=False)['Daily_screen_time_hours'].mean()
```

F3.

```
category_usage_gender =
df_rmv_col.groupby('Gender')[['Social_media_usage_hours',
'Gaming_app_usage_hours', 'Productivity_app_usage_hours']].mean()
```

F3.

```
usage_columns = ['Social_media_usage_hours', 'Productivity_app_usage_hours',
'Gaming_app_usage_hours']
age_group_usage = df_rmv_col.groupby('Age_group')[usage_columns].mean()
```

F4.

```
location_app_usage = df.groupby('Location')[['Social_media_usage_hours',
'Gaming_app_usage_hours', 'Productivity_app_usage_hours']].mean().reset_index()
```

Appendix G. Code for Data Visualization

G1.

```
plt.figure(figsize=(10, 6))
ax = sns.barplot(
    x='Age_group',
    y='Daily_screen_time_hours',
    data=age_group_summary,
    palette='coolwarm',
    errorbar=None
)
for container in ax.containers:
    ax.bar_label(container, fmt='%.2f', label_type='edge', fontsize=9, padding=3)
```

`` plt.title('Daily Screen Time Across Age Groups') plt.xlabel('Age Group') plt.ylabel('Daily Screen Time (Hours)') plt.xticks(rotation=45) plt.show() ```

G2.

```
ax = category_usage_gender.plot(kind='bar', stacked=False, figsize=(10, 6),
color=['#4CAF50', '#2196F3', '#FFC107'])

plt.title('App Category Usage by Gender')
plt.xlabel('Gender')
plt.ylabel('Average Hours of Usage')
plt.legend(title="App Categories", bbox_to_anchor=(1, -0.1), fontsize=9, loc='upper
center')

for container in ax.containers:
    ax.bar_label(container, fmt='%.2f', fontsize=8)

plt.show()
```

G3.

```
ax = age_group_usage.plot(kind='bar', width=0.8, figsize=(12, 6))
plt.title('Average Usage of Social Media, Gaming, and Productivity Apps by Age
Group')
plt.xlabel('Age Group')
plt.ylabel('Average Hours of Usage')
plt.xticks(rotation=45)
plt.legend(title="App Categories", fontsize=9, loc='upper center', bbox_to_anchor=(1,
-0.1))
for container in ax.containers:
    ax.bar_label(container, fmt='%.2f', fontsize=8, padding=3)

plt.show()
```

G4.

```
location_app_usage_melted = location_app_usage.melt(id_vars='Location',
                                    value_vars=['Social_media_usage_hours',
'Gaming_app_usage_hours', 'Productivity_app_usage_hours'],
                                    var_name='App Category',
                                    value_name='Average Usage Hours')
plt.figure(figsize=(14, 8))
ax = sns.barplot(x='Location', y='Average Usage Hours', hue='App Category',
```

```
data=location_app_usage_melted)


for container in ax.containers:
    ax.bar_label(container, fmt='%.2f', fontsize=10, padding=5)
plt.title('App Usage by Location and App Category', fontsize=16)
plt.xlabel('Location', fontsize=12)
plt.ylabel('Average Usage Hours', fontsize=12)
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for readability
plt.legend(title='App Categories', fontsize=9, bbox_to_anchor=(1, -0.1),loc='upper
center')

plt.show()
```