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

df = pd.read_csv(r"C:\Users\ajayv\Downloads\Phone Usage in India\
phone_usage_india.csv")
df.head()
```

```
  User ID  Age  Gender   Location Phone Brand       OS  Screen Time
(hrs/day)  \
0  U00001   53    Male     Mumbai        Vivo  Android
3.7
1  U00002   60   Other      Delhi      Realme      iOS
9.2
2  U00003   37  Female  Ahmedabad       Nokia  Android
4.5
3  U00004   32    Male       Pune     Samsung  Android
11.0
4  U00005   16    Male     Mumbai      Xiaomi      iOS
2.2

   Data Usage (GB/month)  Calls Duration (mins/day)  Number of Apps
Installed  \
0                   23.9                       37.9
104
1                   28.1                       13.7
169
2                   12.3                       66.8
96
3                   25.6                      156.2
146
4                    2.5                      236.2
86

   Social Media Time (hrs/day)  E-commerce Spend (INR/month)  \
0                          3.9                           469
1                          2.8                          4997
2                          3.0                          2381
3                          5.2                          1185
4                          5.5                           106

   Streaming Time (hrs/day)  Gaming Time (hrs/day)  \
0                       5.2                    4.1
1                       5.1                    0.4
2                       1.7                    2.9
3                       3.2                    0.3
4                       3.4                    2.3

   Monthly Recharge Cost (INR)    Primary Use
0                          803      Education
```

```
1                          1526         Gaming
2                          1619  Entertainment
3                          1560  Entertainment
4                           742   Social Media
```

```
df.isnull().sum() #Checking null values
```

```
User ID                          0
Age                              0
Gender                           0
Location                         0
Phone Brand                      0
OS                               0
Screen Time (hrs/day)            0
Data Usage (GB/month)            0
Calls Duration (mins/day)        0
Number of Apps Installed         0
Social Media Time (hrs/day)      0
E-commerce Spend (INR/month)     0
Streaming Time (hrs/day)         0
Gaming Time (hrs/day)            0
Monthly Recharge Cost (INR)      0
Primary Use                      0
dtype: int64
```

```
df.info() #Checking info of data
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17686 entries, 0 to 17685
Data columns (total 16 columns):
 #   Column                          Non-Null Count  Dtype  
---  ------                          --------------  -----  
 0   User ID                         17686 non-null  object 
 1   Age                             17686 non-null  int64  
 2   Gender                          17686 non-null  object 
 3   Location                        17686 non-null  object 
 4   Phone Brand                     17686 non-null  object 
 5   OS                              17686 non-null  object 
 6   Screen Time (hrs/day)           17686 non-null  float64
 7   Data Usage (GB/month)           17686 non-null  float64
 8   Calls Duration (mins/day)       17686 non-null  float64
 9   Number of Apps Installed        17686 non-null  int64  
 10  Social Media Time (hrs/day)     17686 non-null  float64
 11  E-commerce Spend (INR/month)    17686 non-null  int64  
 12  Streaming Time (hrs/day)        17686 non-null  float64
 13  Gaming Time (hrs/day)           17686 non-null  float64
 14  Monthly Recharge Cost (INR)     17686 non-null  int64  
 15  Primary Use                     17686 non-null  object 
dtypes: float64(6), int64(4), object(6)
memory usage: 2.2+ MB
```

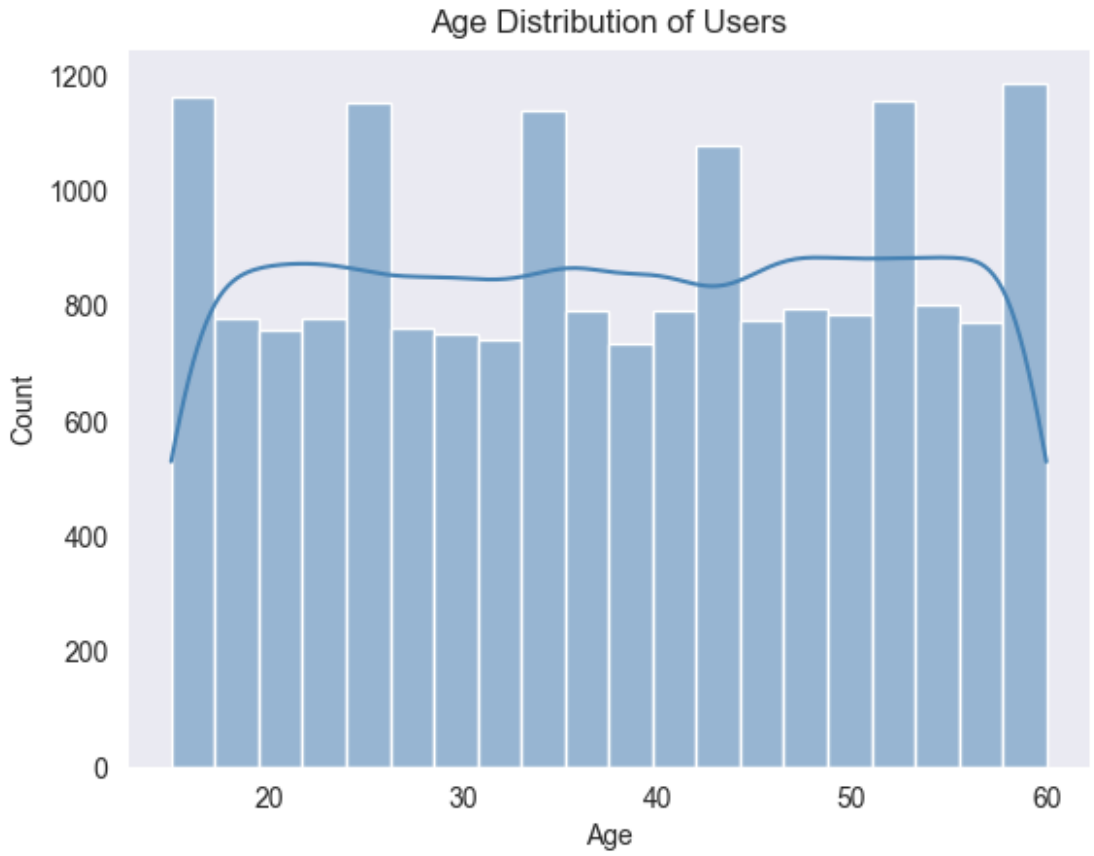
```
df.describe() #statistically summary
```

```
                Age  Screen Time (hrs/day)  Data Usage (GB/month)  \
count  17686.000000           17686.000000           17686.000000
mean      37.584247               6.546376              25.411257
std       13.338252               3.172677              14.122167
min       15.000000               1.000000               1.000000
25%       26.000000               3.800000              13.200000
50%       38.000000               6.600000              25.300000
75%       49.000000               9.300000              37.600000
max       60.000000              12.000000              50.000000

       Calls Duration (mins/day)  Number of Apps Installed  \
count               17686.000000              17686.000000
mean                  151.405846                104.584869
std                    84.923353                 55.217097
min                     5.000000                 10.000000
25%                    77.325000                 57.000000
50%                   150.600000                104.000000
75%                   223.900000                152.000000
max                   300.000000                200.000000

       Social Media Time (hrs/day)  E-commerce Spend (INR/month)  \
count                 17686.000000                  17686.000000
mean                      3.252369                   5075.707848
std                       1.590223                   2871.604841
min                       0.500000                    100.000000
25%                       1.900000                   2587.500000
50%                       3.200000                   5052.000000
75%                       4.600000                   7606.000000
max                       6.000000                  10000.000000

       Streaming Time (hrs/day)  Gaming Time (hrs/day)  \
count              17686.000000           17686.000000
mean                   4.250616               2.490874
std                    2.155683               1.446003
min                    0.500000               0.000000
25%                    2.400000               1.200000
50%                    4.200000               2.500000
75%                    6.100000               3.700000
max                    8.000000               5.000000

       Monthly Recharge Cost (INR)
count                 17686.000000
mean                   1042.785367
std                     552.502067
min                     100.000000
25%                     561.000000
50%                    1040.000000
```

```
75%                          1521.750000
max                          2000.000000
```

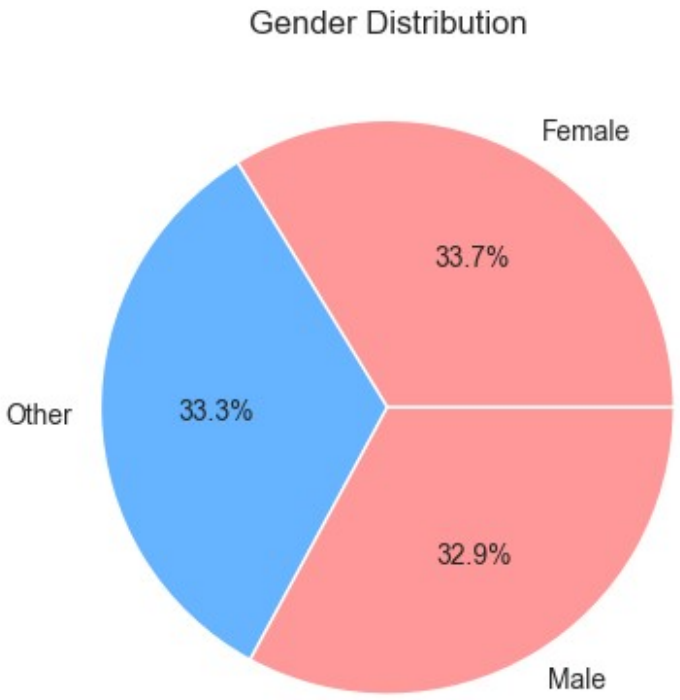
What is the age distribution of users?

```
sns.histplot(df['Age'], bins=20, kde=True, color="steelblue")
plt.title("Age Distribution of Users")
plt.show()
```



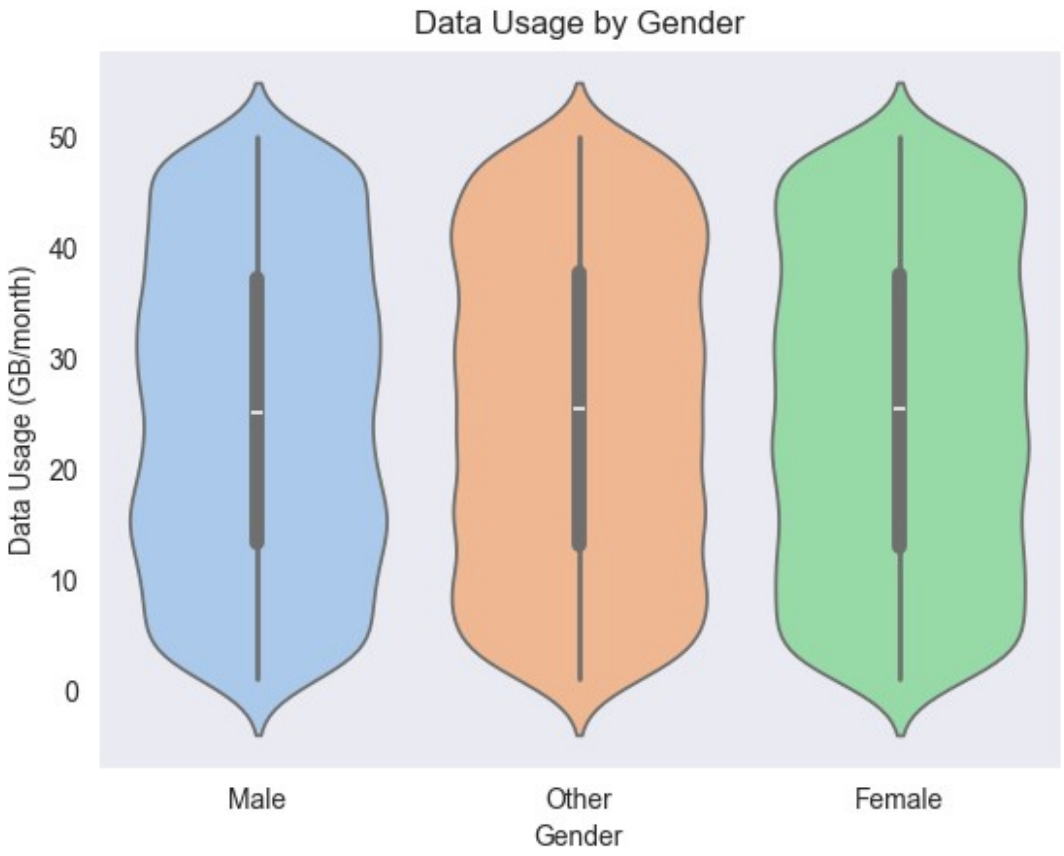
What is the gender distribution in the dataset?

```
df['Gender'].value_counts().plot.pie(autopct="%1.1f%%",
colors=['#ff9999','#66b3ff'])
plt.title("Gender Distribution")
plt.ylabel("")
plt.show()
```



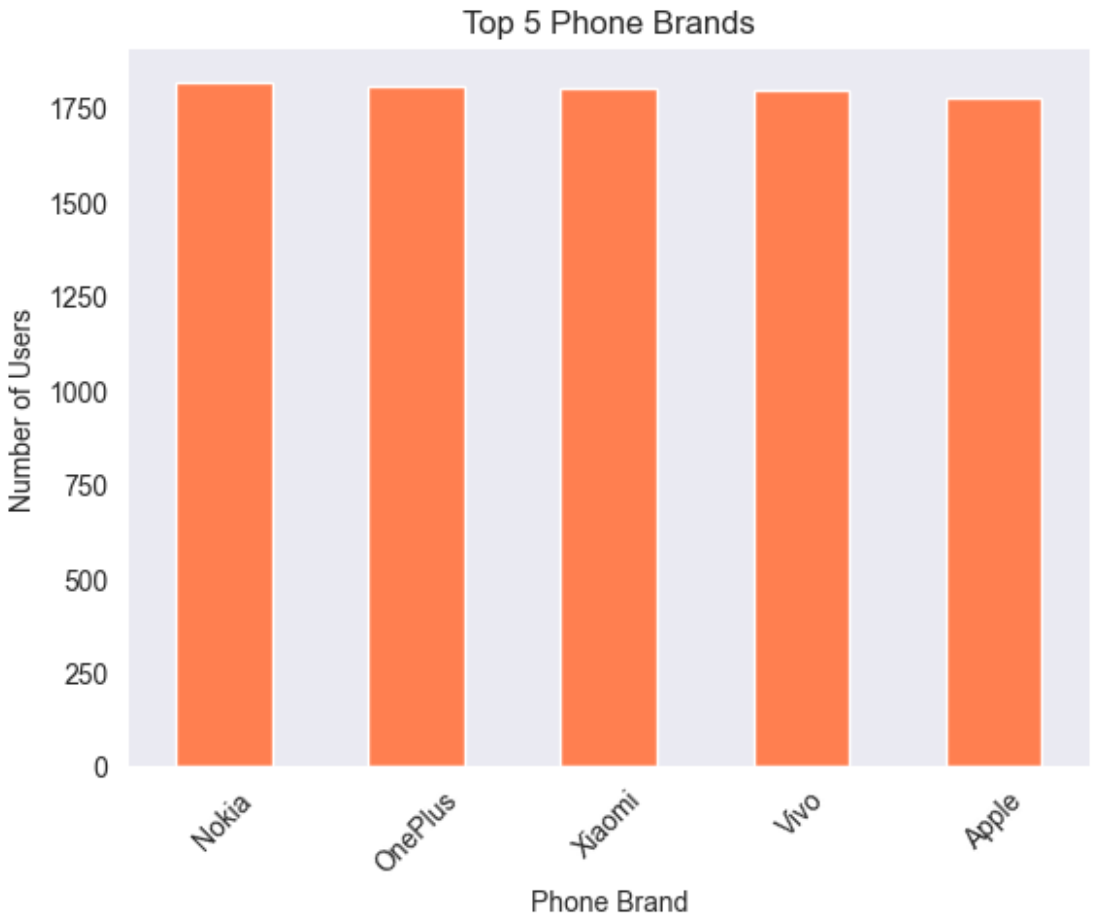
How does data usage vary between genders?

```
sns.violinplot(x="Gender", y="Data Usage (GB/month)", data=df,
palette="pastel", hue="Gender")
plt.title("Data Usage by Gender")
plt.show()
```



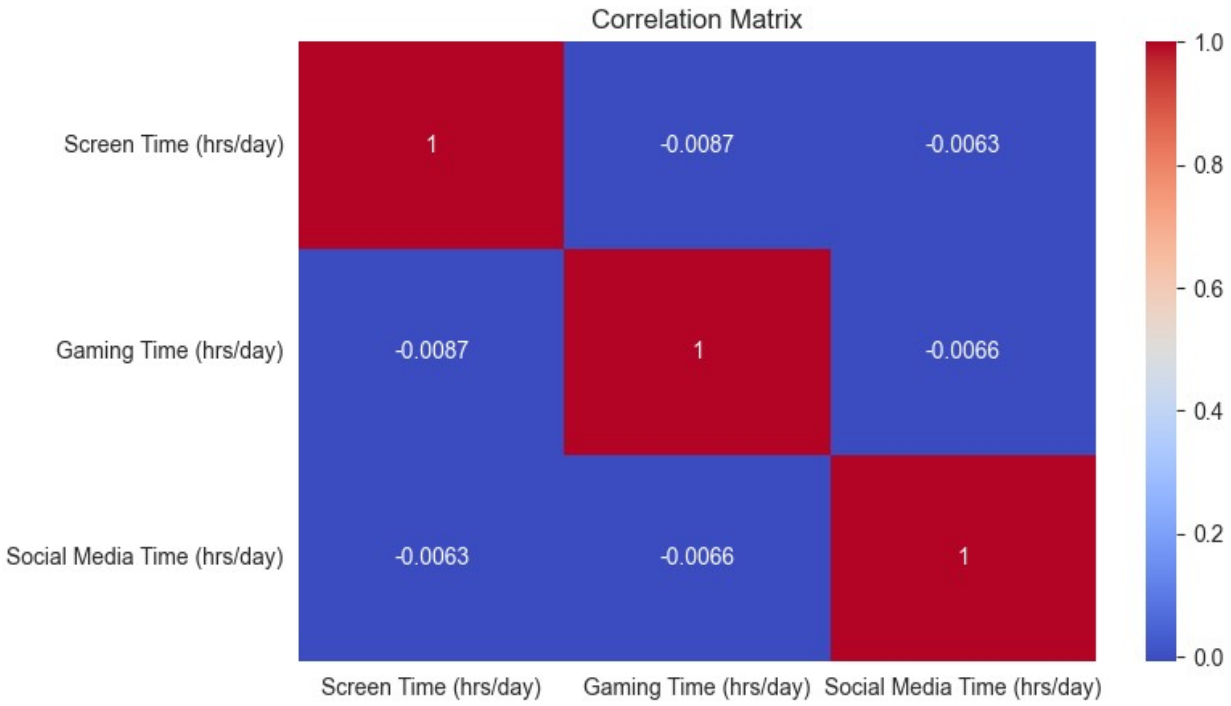
What are the top 5 most used phone brands?

```
df['Phone Brand'].value_counts().head(5).plot(kind='bar',
color="coral")
plt.title("Top 5 Phone Brands")
plt.xlabel("Phone Brand")
plt.ylabel("Number of Users")
plt.xticks(rotation=45)
plt.show()
```



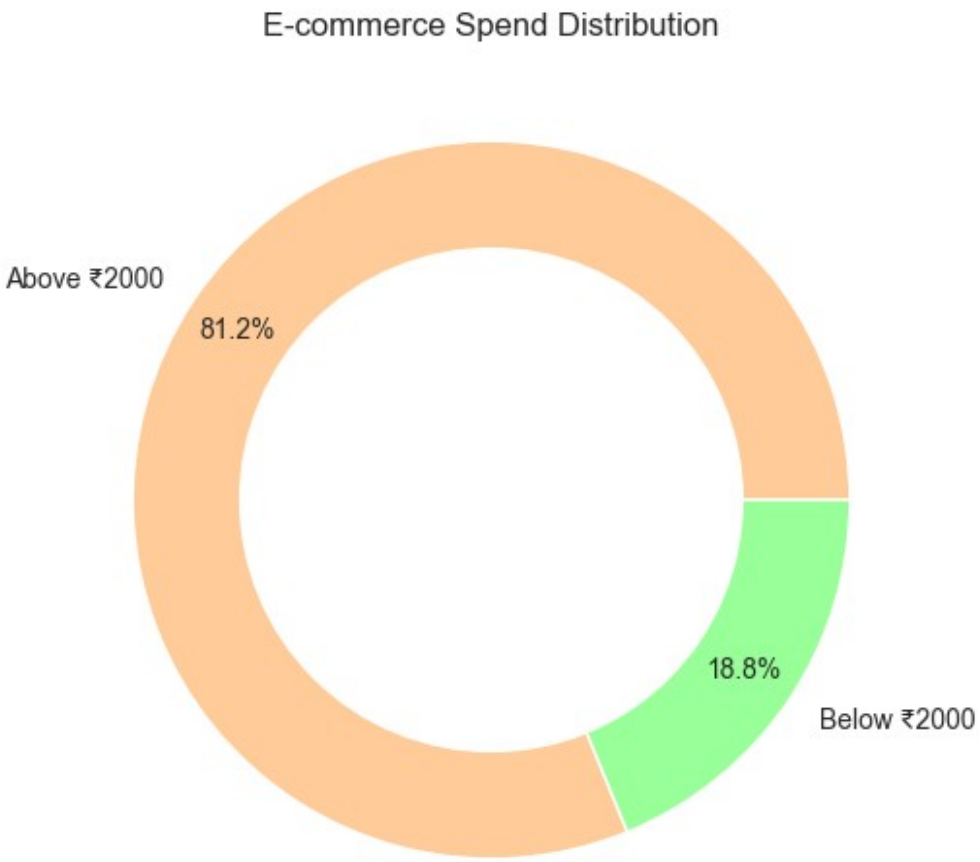
What is the correlation between screen time, gaming time, and social media time?

```
sns.heatmap(df[['Screen Time (hrs/day)', 'Gaming Time (hrs/day)',
'Social Media Time (hrs/day)']].corr(), annot=True, cmap="coolwarm")
plt.title("Correlation Matrix")
plt.show()
```



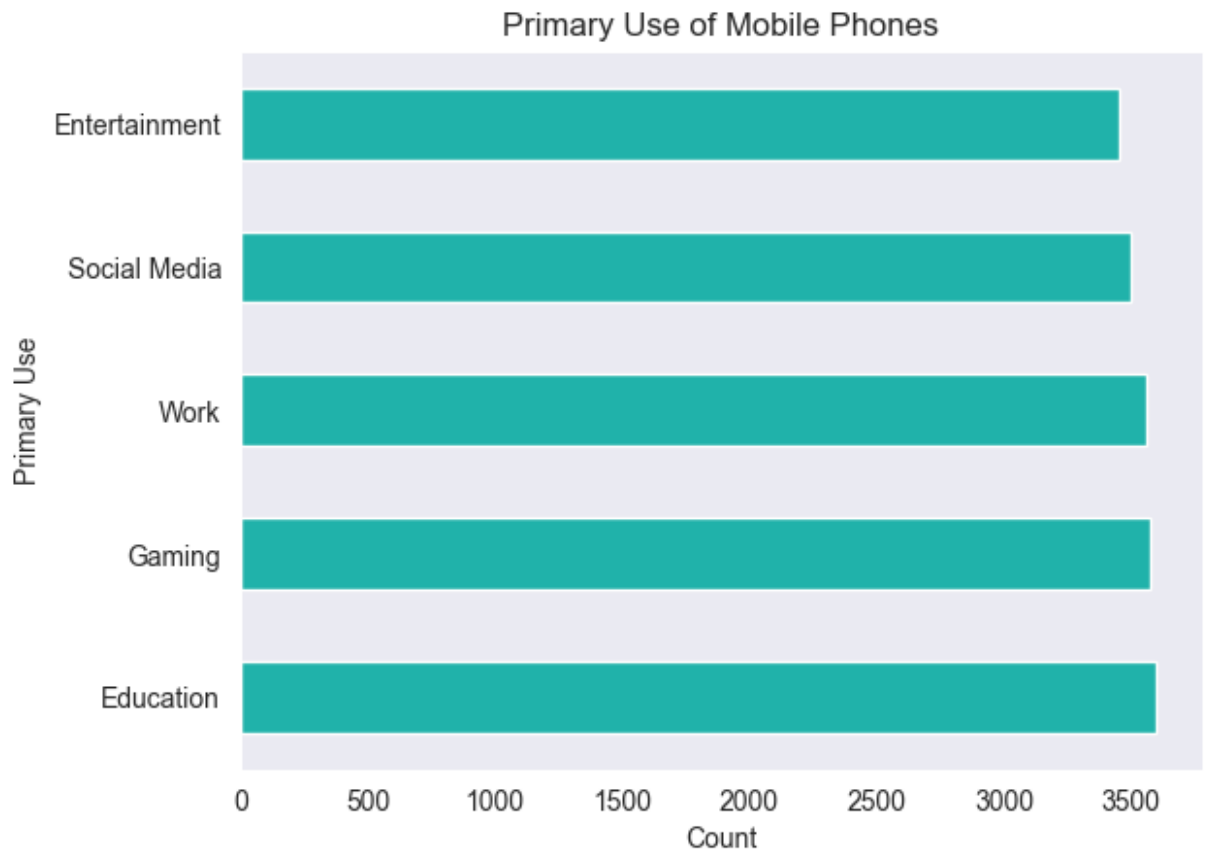
What percentage of users spend more than ₹2000/month on e-commerce?

```
high_spenders = df[df['E-commerce Spend (INR/month)'] > 2000]
low_spenders = df[df['E-commerce Spend (INR/month)'] <= 2000]

sizes = [len(high_spenders), len(low_spenders)]
labels = ['Above ₹2000', 'Below ₹2000']
colors = ['#ffcc99','#99ff99']

plt.figure(figsize=(6,6))
plt.pie(sizes, labels=labels, autopct='%1.1f%%', colors=colors,
pctdistance=0.85)
plt.title("E-commerce Spend Distribution")

# Create a circle at the center for donut shape
center_circle = plt.Circle((0,0),0.70,fc='white')
plt.gcf().gca().add_artist(center_circle)

plt.show()
```



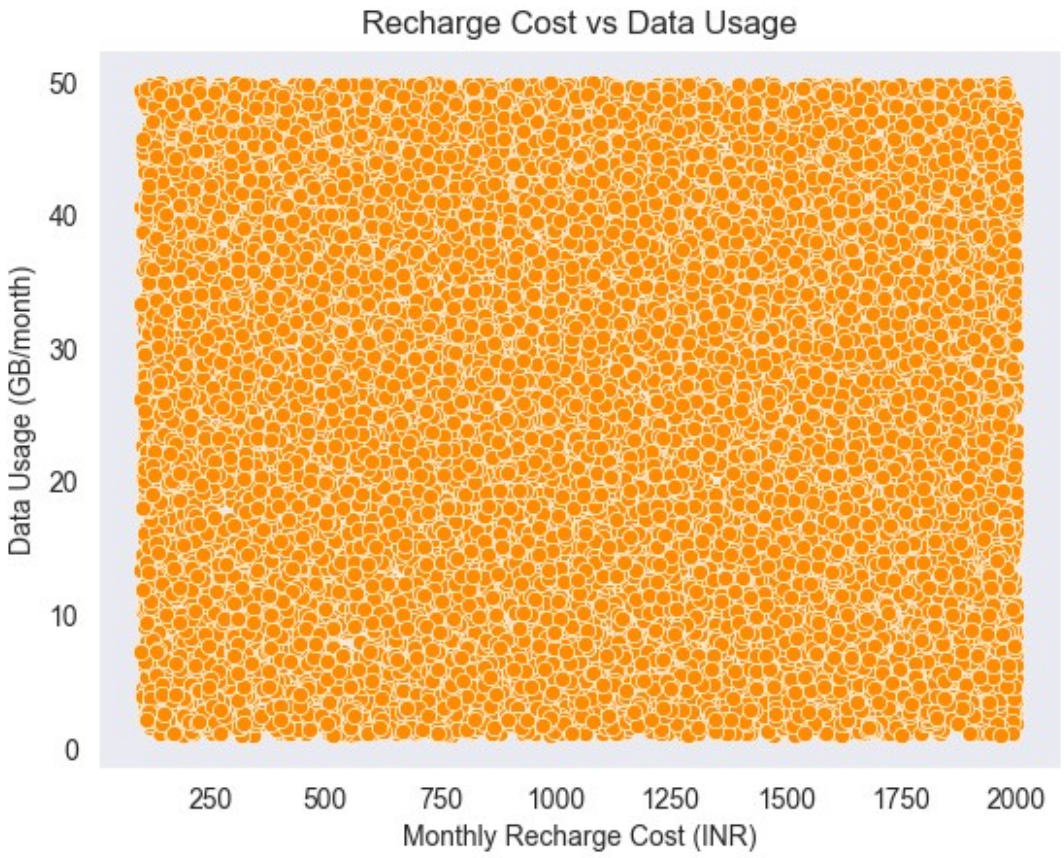
What is the most common primary use of mobile phones?

```
df['Primary Use'].value_counts().plot(kind='barh',
color="lightseagreen")
plt.title("Primary Use of Mobile Phones")
plt.xlabel("Count")
plt.ylabel("Primary Use")
plt.show()
```



What is the relationship between monthly recharge cost and data usage?

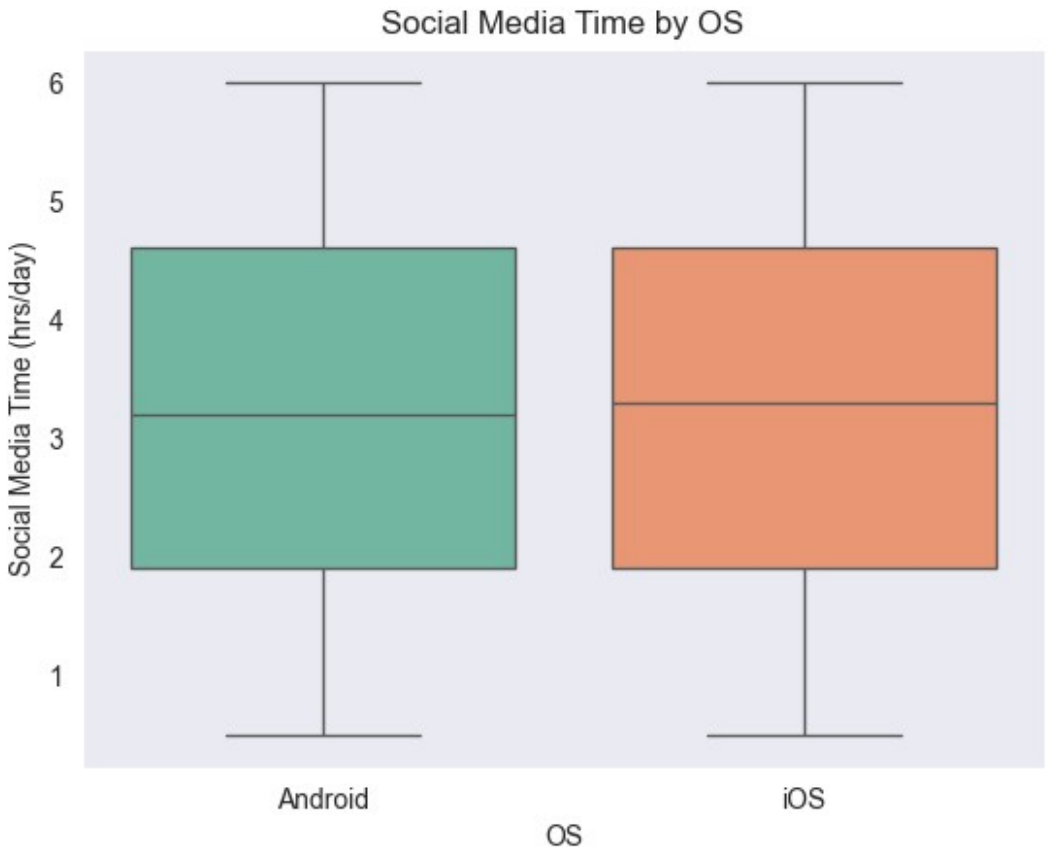
```
sns.scatterplot(x="Monthly Recharge Cost (INR)", y="Data Usage
(GB/month)", data=df, color="darkorange")
plt.title("Recharge Cost vs Data Usage")
plt.xlabel("Monthly Recharge Cost (INR)")
plt.ylabel("Data Usage (GB/month)")
plt.show()
```



Which OS (Android/iOS) has higher social media time?

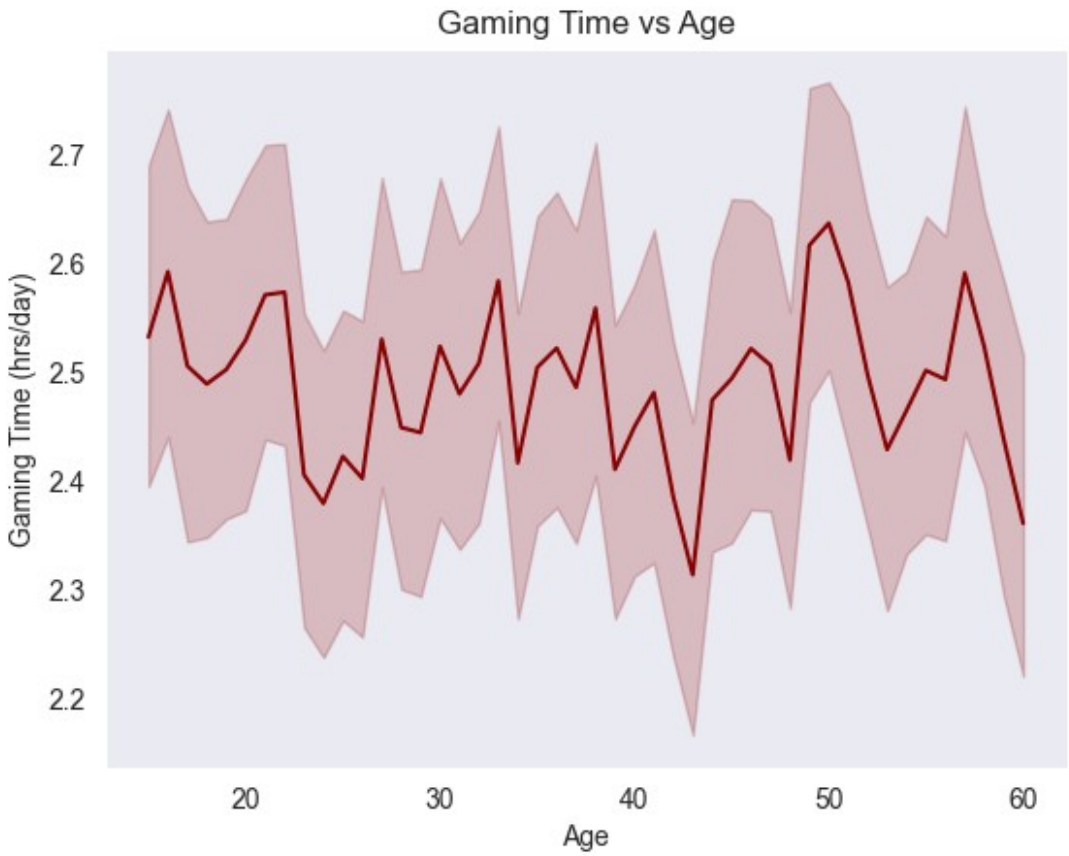
```
sns.boxplot(x="OS", y="Social Media Time (hrs/day)", data=df,
palette="Set2")
plt.title("Social Media Time by OS")
plt.xlabel("OS")
plt.ylabel("Social Media Time (hrs/day)")
plt.show()
```

```
C:\Users\ajayv\AppData\Local\Temp\ipykernel_13656\3188073960.py:1:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be
removed in v0.14.0. Assign the `x` variable to `hue` and set
`legend=False` for the same effect.

  sns.boxplot(x="OS", y="Social Media Time (hrs/day)", data=df,
palette="Set2")
```



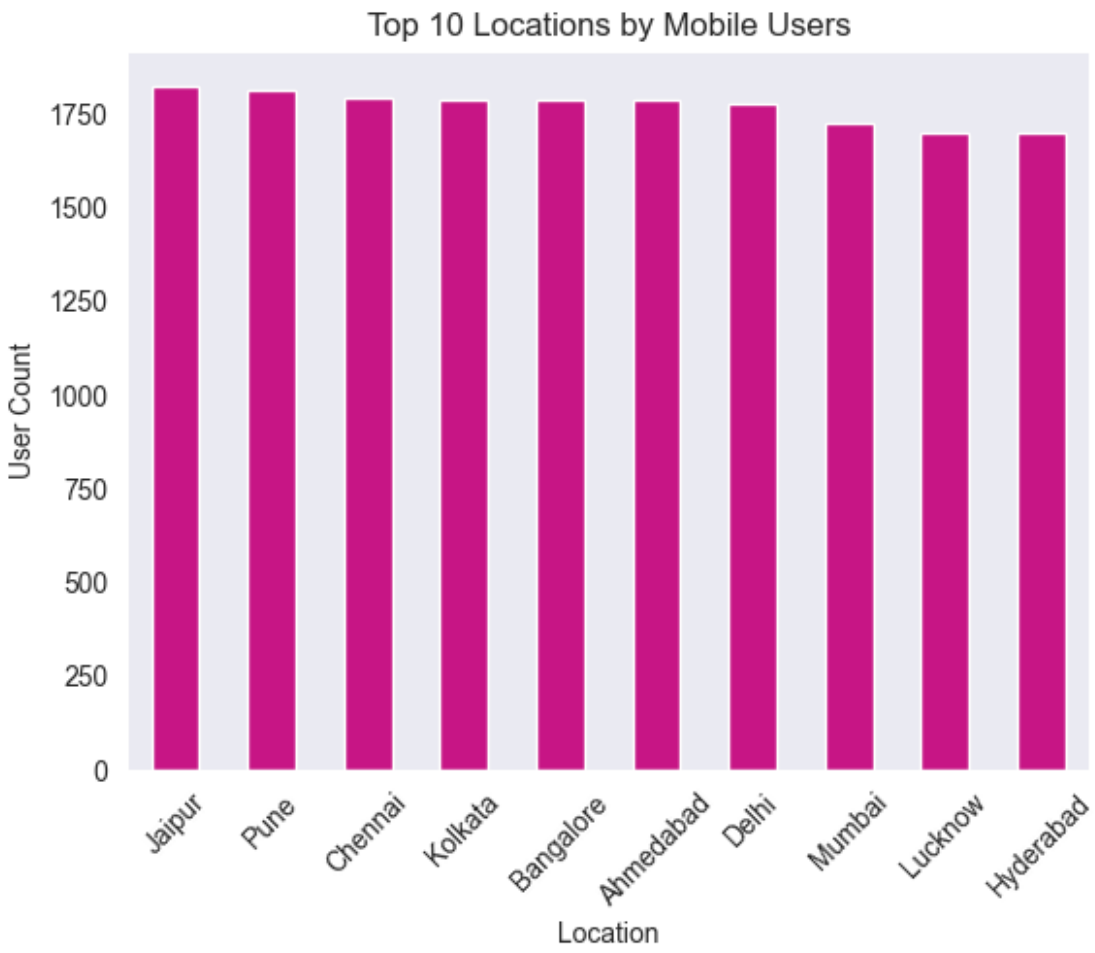
What is the relationship between gaming time and age?

```
sns.lineplot(x="Age", y="Gaming Time (hrs/day)", data=df,
color="darkred")
plt.title("Gaming Time vs Age")
plt.xlabel("Age")
plt.ylabel("Gaming Time (hrs/day)")
plt.show()
```



How does location affect mobile usage?

```
df['Location'].value_counts().head(10).plot(kind='bar',
color="mediumvioletred")
plt.title("Top 10 Locations by Mobile Users")
plt.xlabel("Location")
plt.ylabel("User Count")
plt.xticks(rotation=45)
plt.show()
```


Top 10 Locations by Mobile Users
User Count
1750
1500
1250
1000
750
500
250
0
Jaipur
Pune
Chennai
Kolkata
Bangalore
Ahmedabad
Delhi
Mumbai
Lucknow
Hyderabad
Location