PROJECT NAME : Phone Usage India

Project Members : 1. Prajwal Kore 2. Tanush Bobde

Import CSV file in Google collab

```
from google.colab import files  # to load files------------
upload = files.upload()
```

Choose Files phone_usage_india.csv

- **phone_usage_india.csv**(text/csv) - 1532919 bytes, last modified: 1/13/2025 - 100% done

```
Saving phone_usage_india.csv to phone_usage_india.csv
```

Read CSV File

```
import pandas as pd

df=pd.read_csv('phone_usage_india.csv') #read csv file
print(df.to_string())
```

```
       User ID  Age  Gender   Location   Phone Brand       OS  Screen Time (hrs/day)  Data Usage (GB/month)  Calls Duration (mins/day)
0       U00001   53    Male     Mumbai          Vivo  Android                    3.7                   23.9                       37.9
1       U00002   60   Other      Delhi        Realme      iOS                    9.2                   28.1                       13.7
2       U00003   37  Female  Ahmedabad         Nokia  Android                    4.5                   12.3                       66.8
3       U00004   32    Male       Pune       Samsung  Android                   11.0                   25.6                      156.2
4       U00005   16    Male     Mumbai        Xiaomi      iOS                    2.2                    2.5                      236.2
5       U00006   21    Male     Jaipur          Oppo      iOS                    5.4                   10.6                      210.6
6       U00007   57  Female    Lucknow         Apple      iOS                    6.0                   35.2                      154.5
7       U00008   56   Other    Kolkata        Realme      iOS                    3.1                   43.5                      125.3
8       U00009   46  Female    Kolkata          Oppo  Android                    5.3                   46.4                       21.3
9       U00010   44   Other    Kolkata         Apple      iOS                    9.9                   10.6                      180.2
10      U00011   55   Other    Lucknow         Nokia  Android                    1.6                   23.7                      192.0
11      U00012   41  Female      Delhi          Oppo  Android                    7.5                   23.5                       84.9
12      U00013   53    Male  Bangalore        Realme      iOS                   10.5                    1.4                       33.7
13      U00014   35    Male     Jaipur        Realme  Android                   10.3                   32.4                      247.1
14      U00015   33  Female     Jaipur       Samsung      iOS                    1.7                   33.2                      208.8
15      U00016   52    Male  Bangalore  Google Pixel  Android                    5.1                   28.5                      169.2
16      U00017   46   Other     Mumbai         Apple  Android                    3.1                   23.6                      153.9
17      U00018   54    Male    Kolkata         Apple  Android                   11.6                   12.6                      210.4
18      U00019   50  Female    Chennai  Google Pixel  Android                    5.3                    5.2                      235.6
19      U00020   40   Other  Hyderabad      Motorola      iOS                    9.0                   38.9                      289.6
20      U00021   60    Male       Pune       Samsung      iOS                   10.4                   46.0                       38.0
21      U00022   59   Other       Pune        Xiaomi  Android                   11.7                    9.1                      203.9
22      U00023   46   Other    Chennai       Samsung  Android                    8.1                   15.5                      206.8
23      U00024   16  Female  Hyderabad          Vivo  Android                    1.4                   25.5                      222.0
24      U00025   29  Female  Bangalore        Realme  Android                    7.0                   41.7                       51.0
25      U00026   36    Male  Hyderabad          Vivo      iOS                    5.6                   47.4                      155.4
26      U00027   15    Male      Delhi          Vivo  Android                    9.2                   42.7                      230.9
27      U00028   41   Other     Jaipur        Realme  Android                    8.1                   44.1                       81.8
28      U00029   51  Female    Kolkata          Vivo      iOS                   10.7                   45.1                       94.0
29      U00030   45  Female  Hyderabad        Realme  Android                    6.3                   34.5                      244.3
30      U00031   39  Female     Jaipur      Motorola  Android                    4.0                    9.4                      194.2
31      U00032   40  Female    Lucknow       OnePlus  Android                    8.4                   25.7                      262.8
32      U00033   28  Female    Chennai        Xiaomi      iOS                    5.3                   16.8                      264.9
33      U00034   23  Female    Chennai      Motorola  Android                   11.4                   48.4                      216.8
34      U00035   57    Male      Delhi          Vivo      iOS                    4.1                   31.2                       61.5
35      U00036   23  Female  Ahmedabad          Vivo      iOS                   10.7                   29.0                        9.0
36      U00037   56   Other    Kolkata         Nokia  Android                    9.4                   42.1                      246.7
37      U00038   20   Other    Chennai       Samsung      iOS                   11.9                   45.5                      195.8
38      U00039   22   Other    Kolkata  Google Pixel  Android                    1.6                   34.1                        6.0
39      U00040   19    Male    Kolkata         Apple      iOS                    5.5                    2.9                      212.5
40      U00041   49   Other    Lucknow       Samsung      iOS                    5.2                   43.4                      151.8
41      U00042   54    Male  Ahmedabad          Vivo      iOS                    2.2                   22.0                       30.8
42      U00043   40    Male  Bangalore          Vivo      iOS                    2.5                   10.6                      219.1
43      U00044   52  Female    Kolkata        Xiaomi      iOS                    4.4                   33.8                      189.3
44      U00045   42   Other     Mumbai  Google Pixel      iOS                    4.2                    3.4                      171.6
45      U00046   23   Other  Hyderabad       Samsung      iOS                    4.3                   26.4                      210.5
46      U00047   51   Other     Mumbai       Samsung      iOS                   10.7                   17.6                      293.4
47      U00048   19    Male  Bangalore        Realme      iOS                    3.1                   24.8                       82.1
48      U00049   52   Other      Delhi        Xiaomi  Android                    1.5                   41.0                       95.5
49      U00050   17   Other    Lucknow         Apple      iOS                    4.7                   30.4                      157.7
50      U00051   31  Female  Ahmedabad         Nokia      iOS                   10.4                   41.1                       32.5
51      U00052   49   Other     Jaipur       OnePlus  Android                    8.6                   46.0                      181.8
52      U00053   21    Male      Delhi      Motorola      iOS                    9.4                   15.4                       10.6
53      U00054   51    Male    Lucknow       OnePlus      iOS                    5.4                   15.8                      115.7
54      U00055   32   Other     Mumbai          Vivo      iOS                    5.8                   27.5                       30.0
55      U00056   43   Other  Bangalore        Xiaomi      iOS                    8.6                   14.4                      275.5
```

```
print(df.info()) # check information
```

```
<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
None
```

```
print(df.head()) # first 5 rows
```

```
  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.drop_duplicates(inplace=True)  # Remove duplicate rows
```

```
df.isnull().sum()  # Count missing values per column
```

	0
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['User ID'].unique()  # Check unique values in a categorical column
```

```
array(['U00001', 'U00002', 'U00003', ..., 'U17684', 'U17685', 'U17686'],
      dtype=object)
```

```
df['Screen Time (hrs/day)'] = pd.to_datetime(df['Screen Time (hrs/day)'])
df['Social Media Time (hrs/day)'] = pd.to_datetime(df['Social Media Time (hrs/day)'])
df['Streaming Time (hrs/day)'] = pd.to_datetime(df['Streaming Time (hrs/day)'])
df['Gaming Time (hrs/day)'] = pd.to_datetime(df['Gaming Time (hrs/day)'])
```

```
print(df.info()) # check above all are data type changed
```

```
<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  datetime64[ns]
 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  datetime64[ns]
 11  E-commerce Spend (INR/month)  17686 non-null  int64
 12  Streaming Time (hrs/day)      17686 non-null  datetime64[ns]
 13  Gaming Time (hrs/day)         17686 non-null  datetime64[ns]
 14  Monthly Recharge Cost (INR)   17686 non-null  int64
 15  Primary Use                   17686 non-null  object
dtypes: datetime64[ns](4), float64(2), int64(4), object(6)
memory usage: 2.2+ MB
None
```

```
print(df.dtypes) # Cheek datatypes
```

```
User ID                              object
Age                                   int64
Gender                               object
```

```
Location                              object
Phone Brand                           object
OS                                    object
Screen Time (hrs/day)          datetime64[ns]
Data Usage (GB/month)                 float64
Calls Duration (mins/day)             float64
Number of Apps Installed                int64
Social Media Time (hrs/day)    datetime64[ns]
E-commerce Spend (INR/month)            int64
Streaming Time (hrs/day)       datetime64[ns]
Gaming Time (hrs/day)          datetime64[ns]
Monthly Recharge Cost (INR)             int64
Primary Use                           object
dtype: object
```

Location And their How Many GB data user in perticular city

```
group_by_Location = df.groupby('Location')['Age'].mean()
print(group_by_Location)
```

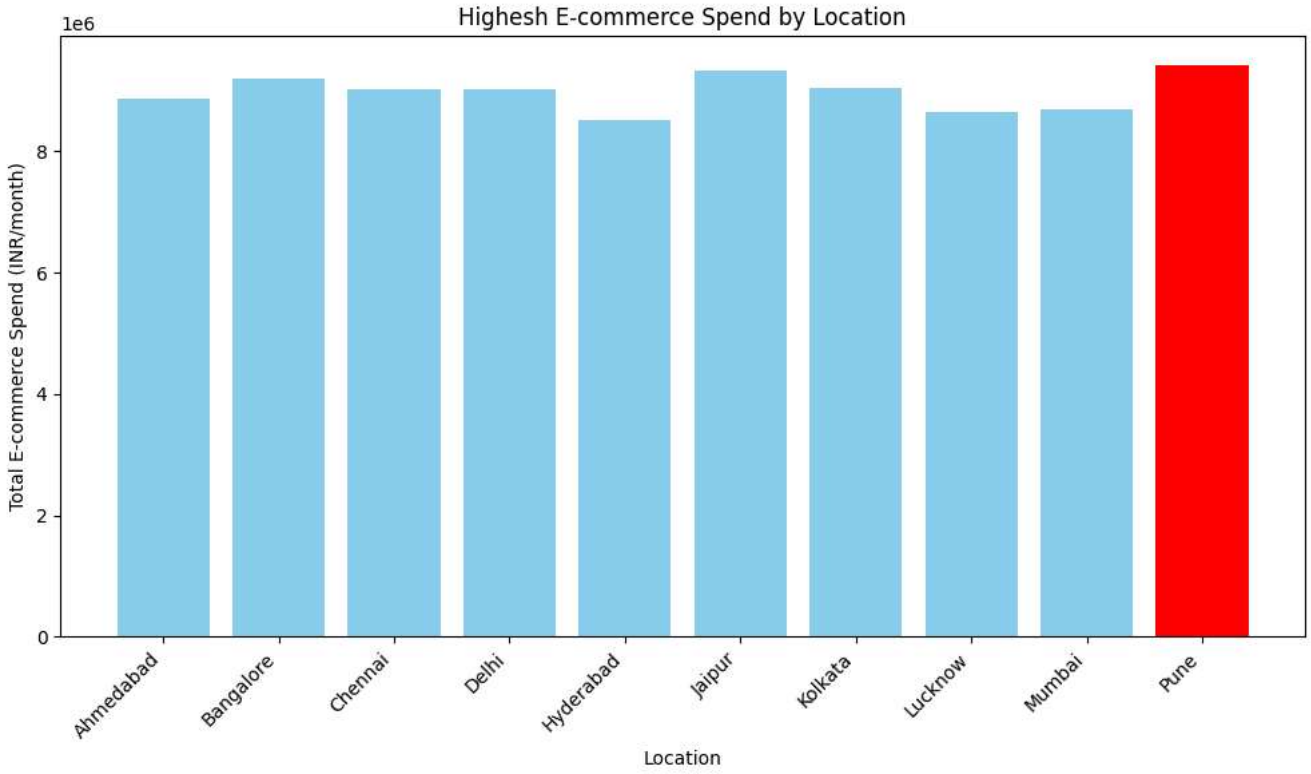
```
Location
Ahmedabad    37.609524
Bangalore    37.208287
Chennai      37.402235
Delhi        37.441127
Hyderabad    37.590588
Jaipur       37.850877
Kolkata      37.678032
Lucknow      38.056471
Mumbai       37.677120
Pune         37.352066
Name: Age, dtype: float64
```

Highest Ecommerce Spend INR/Month by location

```
# Group by location and calculate the total spend for each location
location_spend = df.groupby('Location')['E-commerce Spend (INR/month)'].sum()

# location with the maximum spend
max_spend_location = location_spend.idxmax()

print(f"The location with the maximum E-commerce spend is: {max_spend_location}")
```

```
The location with the maximum E-commerce spend is: Pune
```

```
import matplotlib.pyplot as plt
# bar chart
plt.figure(figsize=(10, 6))
plt.bar(location_spend.index, location_spend.values, color='skyblue')

# Highlight the maximum spend location
plt.bar(max_spend_location, location_spend[max_spend_location], color='red')

plt.xlabel('Location')
plt.ylabel('Total E-commerce Spend (INR/month)')
plt.title('Highesh E-commerce Spend by Location')
plt.xticks(rotation=45, ha='right')
plt.tight_layout()
plt.show()
```

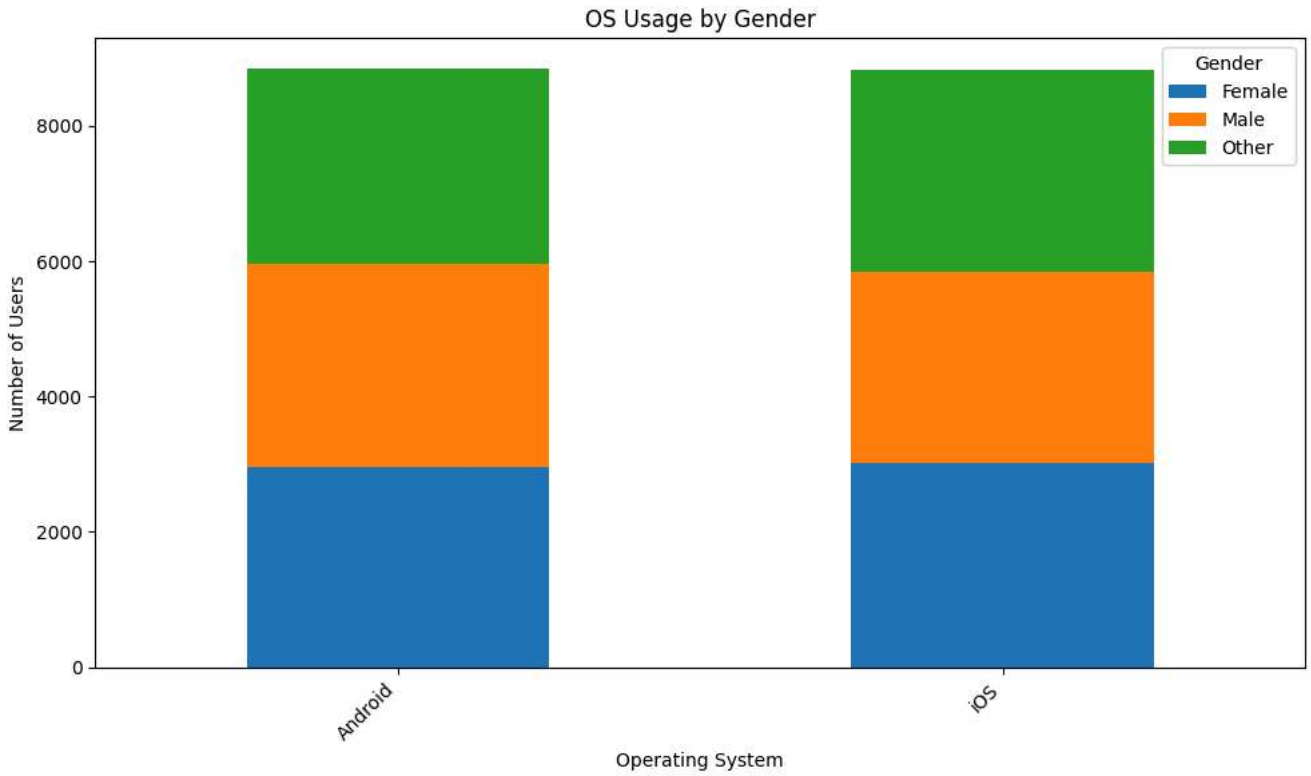


Count users and check Operating systems by genders

```
import pandas as pd

os_gender_grouped = df.groupby(['OS', 'Gender'])['User ID'].count().reset_index()

# Rename the 'User ID' column to 'Count' for better clarity
os_gender_grouped = os_gender_grouped.rename(columns={'User ID': 'Count'})

# Print the grouped data
print(os_gender_grouped)
```

```
         OS  Gender  Count
0  Android  Female   2961
1  Android    Male   2996
2  Android   Other   2894
3      iOS  Female   3008
4      iOS    Male   2829
5      iOS   Other   2998
```

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

# stacked bar chart
os_gender_grouped.pivot(index='OS', columns='Gender', values='Count').plot(kind='bar', stacked=True, figsize=(10, 6))

plt.xlabel('Operating System')
plt.ylabel('Number of Users')
plt.title('OS Usage by Gender')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Gender')
plt.tight_layout()
plt.show()
```

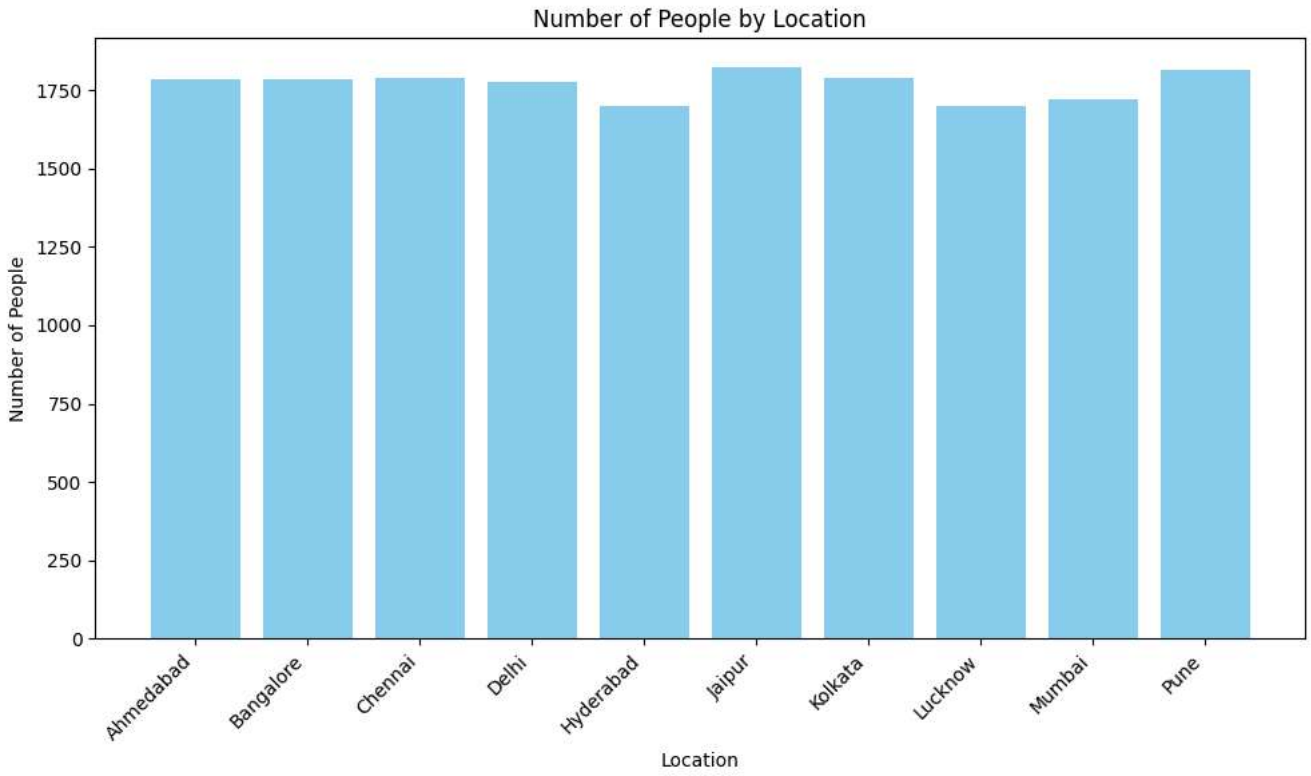


Count Users by location

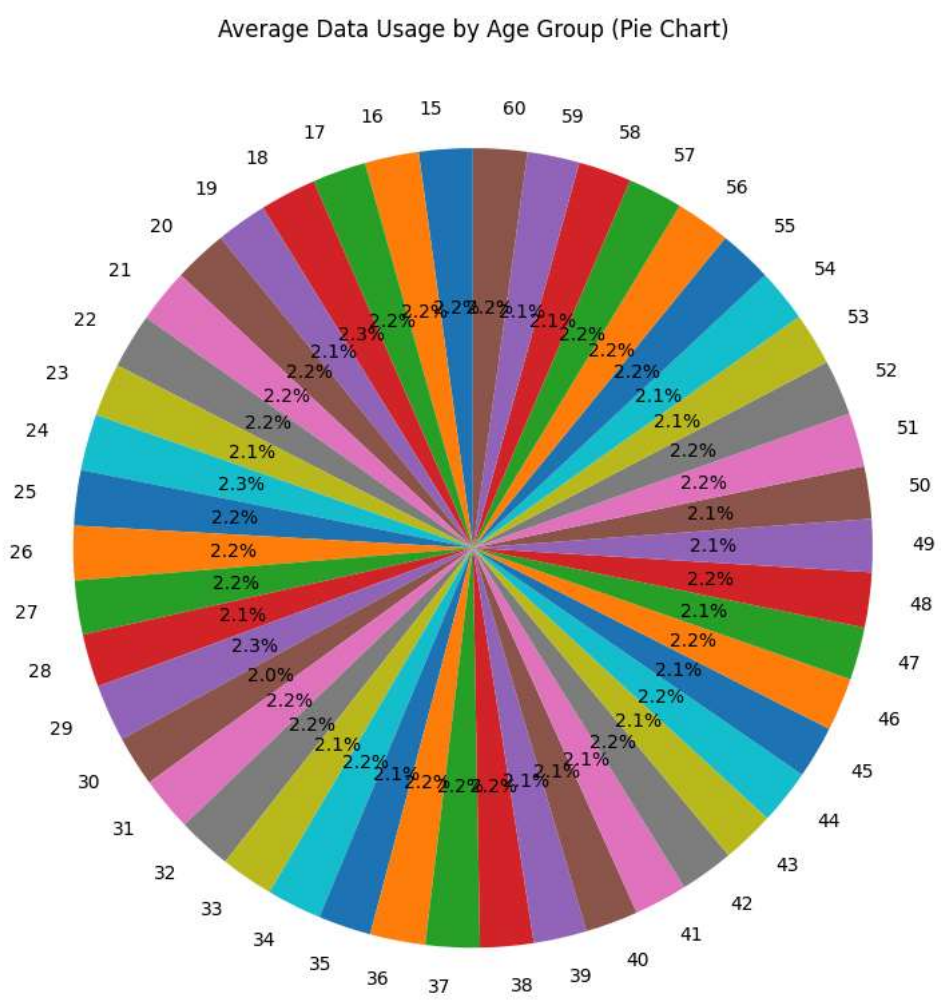
```
# Total users from different location count
import pandas as pd

location_counts = df.groupby('Location')['User ID'].count().reset_index()

# Print the location counts
print(location_counts)
```

```
     Location  User ID
0  Ahmedabad     1785
1  Bangalore     1786
2    Chennai     1790
3      Delhi     1775
4  Hyderabad     1700
5     Jaipur     1824
6    Kolkata     1789
7    Lucknow     1700
8     Mumbai     1722
9       Pune     1815
```

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

# Create a bar chart
plt.figure(figsize=(10, 6))  # Adjust figure size if needed
plt.bar(location_counts['Location'], location_counts['User ID'], color='skyblue')
plt.xlabel('Location')
plt.ylabel('Number of People')
plt.title('Number of People by Location')
plt.xticks(rotation=45, ha='right')  # Rotate x-axis labels for better readability
plt.tight_layout()
plt.show()
```




```
import matplotlib.pyplot as plt

group_by_Location = df.groupby('Age')['Data Usage (GB/month)'].mean()

# pie chart
plt.figure(figsize=(8, 8))
plt.pie(group_by_Location.values, labels=group_by_Location.index, autopct='%1.1f%%', startangle=90)
plt.title('Average Data Usage by Age Group (Pie Chart)')
plt.tight_layout()
plt.show()
```



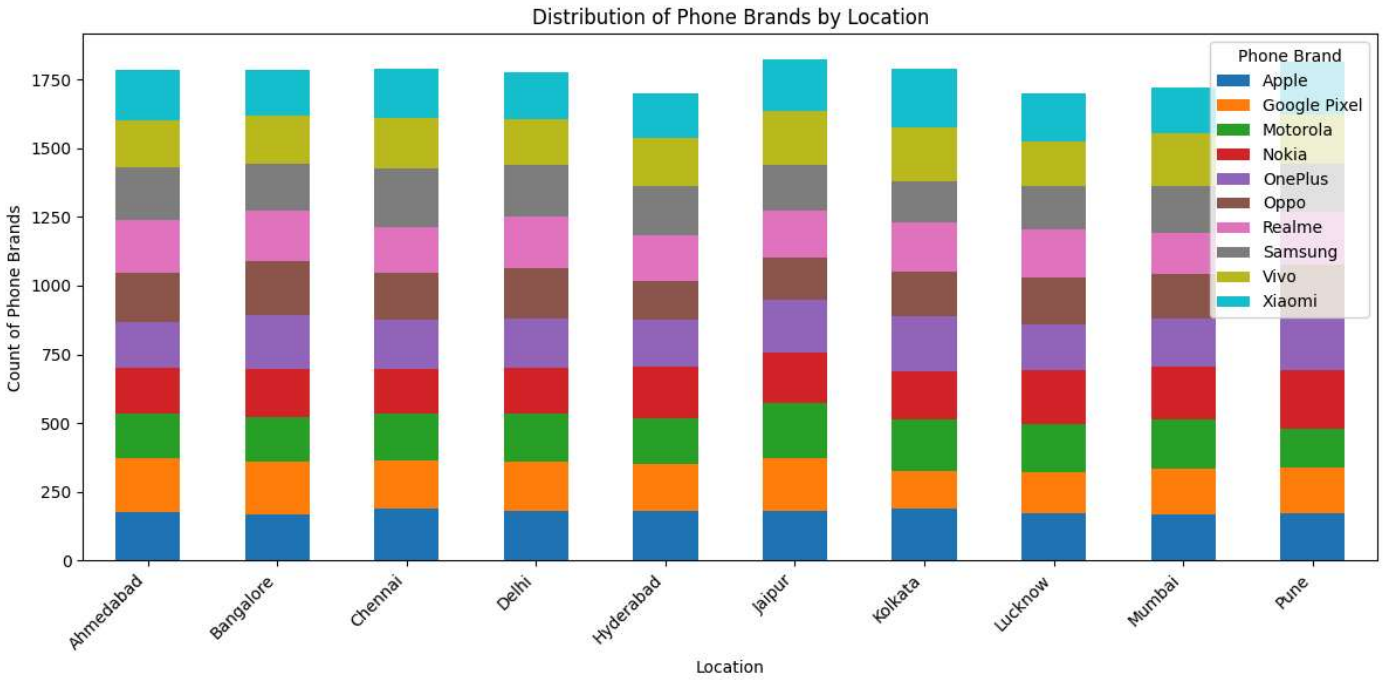
Distribution of Phone Brands by Location

```
group_by_Location = df.groupby('Data Usage (GB/month)')['Age'].max()
print(group_by_Location)
```

```
Data Usage (GB/month)
1.0     60
1.1     58
1.2     59
1.3     59
1.4     57
        ..
49.6    57
49.7    59
49.8    58
49.9    60
50.0    60
Name: Age, Length: 491, dtype: int64
```

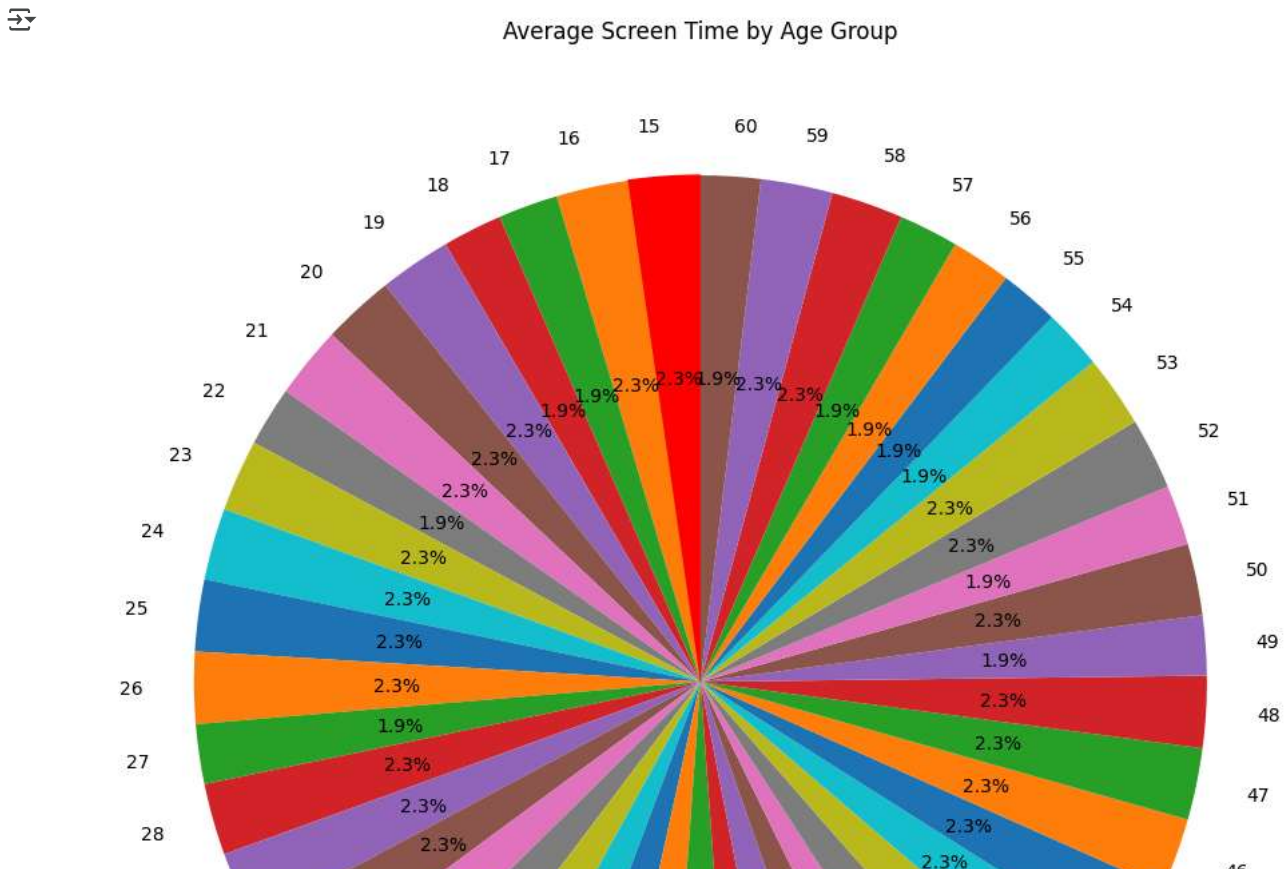
```
import matplotlib.pyplot as plt
import pandas as pd

# Group data by location and phone brand, then count occurrences
grouped_data = df.groupby(['Location', 'Phone Brand'])['Phone Brand'].count().unstack()

# stacked bar chart
grouped_data.plot(kind='bar', stacked=True, figsize=(12, 6))
plt.xlabel('Location')
plt.ylabel('Count of Phone Brands')
plt.title('Distribution of Phone Brands by Location')
plt.xticks(rotation=45, ha='right')
plt.legend(title='Phone Brand')
```

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



Average Screen Time by Age Group

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

# Group data by age and calculate average screen time
age_screen_time = df.groupby('Age')['Screen Time (hrs/day)'].mean()

# age group with the highest average screen time
highest_screen_time_age = age_screen_time.idxmax()

# pie chart
plt.figure(figsize=(15,10))
plt.pie(age_screen_time, labels=age_screen_time.index, autopct='%1.1f%%', startangle=90)
plt.title('Average Screen Time by Age Group')

# Highlight the highest screen time age group
plt.gca().patches[age_screen_time.index.get_loc(highest_screen_time_age)].set_color('red')

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




```
male_data = df[df['Gender'] == 'Male']
female_data = df[df['Gender'] == 'Female']

grouped_by_gender = df.groupby('Gender')
```